

Complex Causal Structures

Extensions of a Regularity Theory of Causation

Michael Baumgartner

Complex Causal Structures
Extensions of a Regularity Theory of Causation

Complex Causal Structures

Extensions of a Regularity Theory of Causation

Michael Baumgartner

2007

Acknowledgements

I would like to thank the people who have contributed to and commented on my work on causation and the ideas contained in this study. Gerd Grasshoff has introduced me to the topic of causation and has shared his valuable insights on the cause-effect relation with me. In countless discussions Timm Lampert has been my highly appreciated sparring partner. In an extraordinarily productive atmosphere he has helped me shape my ideas on virtually every aspect of causation. In addition to Timm Lampert, the development of the procedure of causal reasoning to be proposed in the following has greatly benefited from Daniel Engler's programming skills. Tim Ræz has contributed with serviceable comments on earlier drafts and corrections thereof. Alessa Hool and Rafael Scholl have proofread selected parts of the text.

Many of the ideas contained in this study have evolved in the context of an introductory course on the philosophy of causation that Gerd Grasshoff and I have developed at the University of Bern. Many participants of this course have provided helpful feedback on my answers to the questions addressed in the following. I have also profited from the participants of our research colloquium, where I have been given a chance to report my results repeatedly.

Finally, I would like to express my gratitude to the Swiss National Science Foundation for supporting my work (subsidies 1114-066803.01/1 and 101311-103988/1).

CONTENTS

<i>Acknowledgments</i>	vi
<i>1. Introduction</i>	1
1.1 Why a Regularity Account of Causation?	1
1.2 Common Objections to Regularity Accounts	2
1.2.1 Hume's Legacy	2
1.2.2 Imperfect Regularities	4
1.2.3 Monotony	6
1.2.4 Empty Regularities	8
1.2.5 Non-Symmetry	9
1.2.6 Spurious Regularities	13
1.2.7 Single-Case Regularities	15
1.2.8 Indeterminism	18
1.3 Outline	19
<i>2. Conceptual Fundament</i>	21
2.1 Introduction	21
2.2 The Relata of Causation	21
2.2.1 Fact vs. Event Causation	21
Arguments from the Identity of the Causal Relata	24
The Slingshot-Argument	28
The Argument from Spatiotemporal Locality	41
The Argument from Negation	44
The Argument from the Progressive Localization of Causes	45
Summing Up	49
2.2.2 Singular Events vs. Event Types	50
Singular Events	51
Event Types, Factors	56
2.3 Causal Relevance	65
2.3.1 Singular vs. General Causation	66
Causal Statements	66
Relational Properties	67
2.3.2 Graphical Notation	73
2.3.3 Direct vs. Indirect Relevance	74
2.3.4 Complex Causes and Coincidences	75

2.3.5	Alternative Causes	77
2.3.6	Common Causes	78
2.3.7	Cycles	79
2.4	Causal Principles	80
2.4.1	Principle of Determinism	80
2.4.2	Principle of Causality	81
2.4.3	Principle of Relevance	82
2.4.4	Principle of Persistent Relevance	83
2.5	Features of Causation	84
3.	<i>Causal Relevance Analyzed – Minimal Theories</i>	87
3.1	Introduction	87
3.2	Sufficient and Necessary Conditions in First-Order Logic	90
3.3	Minimally Sufficient Conditions	95
3.4	Minimally Necessary Conditions	98
3.5	Minimal Theories	101
3.6	Causal Relevance Analyzed	106
3.6.1	A First Proposal	106
3.6.2	The Direction of General Causation	111
3.6.3	Multiple Minimal Theories – Specification of Causal Structures	115
3.6.4	Minimal Theories and Negative Factors	118
	Negative Factors as Effects	121
	Negative Factors as Causes	124
3.6.5	Simple vs. Complex Minimal Theories (I)	133
3.6.6	Causal Relevance Defined	136
3.7	Causation Among Events – Singular Causation	138
3.8	Abstraction from Relational Constraints	144
4.	<i>The Chain-Problem</i>	147
4.1	Introduction	147
4.2	Mapping Minimal Theories onto Complex Causal Structures	150
4.3	Chains vs. Entangled Epiphenomena	155
4.3.1	Inference to Complex Causal Structures – The Basic Idea of an Inference Procedure	155
4.3.2	Different Causal Structures, one Coincidence Frame	155
4.4	Causal Intuition and the Distinction Between Chains and Entangled Epiphenomena	160
4.5	Inference to Chains and the Conceptual Fundament of a Regularity Theory	162
4.5.1	Causal Interpretation of Entangled Factors	163
4.5.2	Simple vs. Complex Minimal Theories (II)	170
4.6	The R-Solution?	172
4.6.1	Direction of Time	172

4.6.2	Spatiotemporal Adjacency	174
4.6.3	Energy Transfer	180
4.6.4	Non-transitive R	182
4.6.5	Summing Up	185
4.7	Causal Interpretation of Minimally Necessary Conditions	186
4.8	Causal Reasoning and Coincidence Frames	188
5.	<i>Analysis of Complex Causal Structures</i>	191
5.1	Introduction	191
5.2	Homogeneity	192
5.3	The State of the Art in Causal Reasoning within a Regularity Theoretic Framework	200
5.3.1	Method of Difference	200
5.3.2	The four-field test	202
	Causal Independence of Test-Factors	205
	Inference Rules	207
	Rule of Difference	210
	Rule of Combination	211
	Rule of Alternatives	212
5.3.3	Modularity	213
5.4	Limitations of the Existing Inference Procedure	214
5.5	Coincidence Analysis	217
5.5.1	The Basic Idea	217
5.5.2	Input Data	218
5.5.3	Presuppositions	228
	Exhaustiveness of Coincidence Data	228
	Homogeneity	230
5.5.4	Identification of Potential Effects	233
5.5.5	Identification and Minimalization of Sufficient Conditions	235
5.5.6	Identification and Minimalization of Necessary Conditions	239
5.5.7	Framing Minimal Theories	242
5.5.8	Causal Interpretation	245
5.5.9	Summary of the Algorithm	245
5.6	Further Examples	247
5.7	Non Interpretable Coincidence Lists	252
5.8	Empirical Exhaustiveness Violated	254
5.9	Summary	257
	<i>Appendix</i>	259
	Notation	261
	Definitions	261
	Proofs	264

Index 267

Bibliography 272

1. INTRODUCTION

1.1 *Why a Regularity Account of Causation?*

A mere glance at the abundance of controversial literature on causation published during the past 30 years reveals that regularity accounts of causation – until very recently – virtually vanished from the scene. For lack of space and interest, studies not primarily concerned with causation every now and then roughly explicated our causal intuitions in terms of regularities, but hardly anybody seriously wanting to analyze causation resorted to regularity accounts any more. Problems encountered within other theoretical frameworks have lately induced authors working on causation, laws of nature, or methodologies of causal reasoning – as e.g. May (1999), Ragin (2000), Graßhoff and May (2001), Swartz (2003), Halpin (2003) – to direct their attention back to regularity theoretic analyses. Prompted by these recent reanimations of regularity theories the present study will deviate from the predominant tradition with respect to philosophical analyses of causation. Regardless of the widespread skepticism towards the merits and theoretical prospects of regularity accounts, I shall first present a regularity theory of causation and then develop a procedure of causal reasoning based on that theory.

In view of the unfashionableness of regularity accounts, the choice of this theoretical framework, of course, calls for some justification. The main reason for adopting a regularity theoretic framework stems from its outstanding performance as regards two cardinal aspects of an analysis of causation, both of which will be key issues of the study at hand. First, while the most notable rival theories have given up on attempting to provide a reductive analysis of causation, regularity accounts – properly conceived – offer the promising prospect of explicating the cause-effect relation in entirely non-causal terms. Second, it is one thing to determine under what conditions two single events or factors are causally related, but a wholly different matter to assemble single cause-effect dependencies to complex causal structures. The regularity theoretic procedure of causal reasoning developed in the following exceeds existing alternatives with respect to essential aspects of causal structuring, such as gathering parts of complex causes, identifying alternative causes, and distinguishing among interwoven layers of causal structures. In short, the study at hand provides a regularity theoretic analysis of causation, because, contrary to the hastily declared overall failure of regularity theories, the latter provide valuable analytical means not available within other theoretical frameworks.

Apart from these as yet to be substantiated merits, there are several commonly acknowledged advantages of an analysis of causation in terms of regularities. A regularity theoretic notion of causation directly mirrors central pre-theoretic intuitions expressible in well-known principles as “The same cause is always accompanied by the same effect” or “If no cause is present, no effect occurs”. Moreover, the conceptual apparatus resorted to by the analysans of a regularity theoretic notion of causation is fully embedded within the uncontroversial and well mastered area of extensional standard logic. Furthermore, unlike e.g. counterfactual accounts, regularity theories straightforwardly handle cases of overdetermination and preemption.¹ As against interventionist or manipulatory accounts, analyses of causation in terms of regularities do not run the risk of being anthropocentric, i.e. applicable to the domain of human action only. Contrary to probabilistic accounts, regularity theories are not compromised by paradoxical data as, for instance, generated in cases of Simpson’s Paradox. Finally, while methodologies of causal reasoning embedded in the framework of probabilistic causation only produce informative outputs provided that analyzed conditional probabilities are lower than 1, i.e. provided that causes do not in a strict sense determine their effects, a regularity theoretic framework is custom-built to deterministic causal dependencies and provides the means to properly uncover such dependencies.

Prior to presenting a regularity theory that does justice to the above claims, the venture of professing an account of causation which, for the past 30 years, has been claimed to be futile must be defended against its many critics. A detailed and full defence, however, presupposes the results of the study to be undertaken here and, hence, cannot be satisfactorily provided in these introductory remarks. Apart from furnishing an outline of the inquiry to come, this introductory chapter will thence simply review the objections that have traditionally been raised against regularity accounts and, in a nutshell, sketch ways to refute them. The sketches will then be gradually fleshed out as we proceed.

1.2 Common Objections to Regularity Accounts

1.2.1 Hume’s Legacy

The philosophical core of regularity theories of causation, concisely put, consists of three main tenets: (i) the causal relation is not an ontological primitive, (ii) general causation² – causation on type level – is the primary analysandum, and (iii) universal regularities among event types or *factors* constitute the primary analysans.

¹ In fact, regularity accounts are sometimes criticized for not adequately handling cases of preemption (cf. e.g. Collins, Hall, and Paul (2004), ch. 1, §3). However, these objections target regularity accounts whose primary causal relata are events – as in case of Mackie’s (1963) – and not event types – as in case of Mackie’s (1974). The latter variant of regularity analyses, to which the account presented here pertains as well, are not affected by the mentioned objections (cf. Graßhoff and May (2001), pp. 104-105).

² There are some regularity theoretic proposals that do not subscribe to this tenet, e.g. Mackie (1965). Some criticism raised against regularity theories over the past three decades targets this kind of singularist account (cf. Collins, Hall, and Paul (2004) or Kim (1973)). As the study at hand is only

There are – at least – two causal relations, one on type and another on token level. “Drinking is a cause of drunkenness” is a case of *general causation*, i.e. causation among factors, while “Shamus’ drinking of 6 beers at noon on September 7, 2004 causes Shamus’ drunkenness in the afternoon of September 7, 2004” relates token events and, accordingly, is a case of *singular causation*.³ A factor that is related to another factor in terms of general causation is said to be *causally relevant* to the latter. For brevity, if a first factor is causally relevant to a second factor, the first can also be referred to as a *cause* of the second. Moreover, such as to avoid unnecessary terminological complications, we shall often simply speak of *causation* whenever both causal relations are at issue or whenever the context clarifies whether singular or general causation is under consideration. Factors are taken to be similarity sets of event tokens. They can be seen as sets of type identical token events, of events that share at least one feature. While token events are spatiotemporally located particulars, event types are generic entities. Whenever a member of a similarity set that corresponds to an event type occurs, the latter is said to be *instantiated*.⁴

According to Hume, the godfather of regularity theories, single event sequences are not identifiable as being of causal nature by some inherent physical feature or property. A causal interpretation of an event sequence is warranted only if the corresponding events⁵ instantiate factors one of which is related in terms of general causation to the other. Thus, as singular causation can be straightforwardly accounted for given an analysis of general causation, spelling out the latter relation constitutes the core goal of a regularity theory – hence tenet (ii). Hume famously puts it as follows:

It appears, that, in single instances of the operation of bodies, we never can, by our utmost scrutiny, discover any thing but one event following another; without being able to comprehend any force or power, by which the cause operates, or any connexion between it and its supposed effect. (...) One event follows another; but we never can observe any tie between them.⁶

(...) we may define a cause to be an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second.⁷

Hume takes factors to be related in terms of general causation or causal relevance if they satisfy a conditional as “Whenever *A* is instantiated, *B* is instantiated” such

concerned with regularity theories whose primary analysandum is general causation, these singularist objections will be disregarded in the present context (cf. section 2.3.1).

³ For more details on the distinction between singular and general causation cf. section 2.3.1.

⁴ The notions of an event and of a factor or event type will be properly spelled out in chapter 2.

⁵ This focus on events does not straightforwardly cover cases of causally related absences or omissions. The problems posed by causal dependencies as between omitted vaccination and contracting influenza will be neglected in the present context. They are treated in section 3.6.4. For interesting proposals on how to deal with causation among absences cf. also Collins, Hall, and Paul (2004).

⁶ Hume (1999 (1748)), p. 144.

⁷ Hume (1999 (1748)), p. 146.

that the instances of A and B differ and are spatiotemporally proximate.⁸ Events do not cause themselves – no self-causation – and effects occur nearby their causes – no action at a distance. For an event a to be identified as a token level cause of another event b , it is, according to this conception, required that a instantiates a factor A whose instances are always followed by events of type B , which is instantiated by b . Type level causes are thus analyzed to be *sufficient conditions* of their type level effects.⁹ This yields a first Hume-inspired proposal for a regularity theoretic account of causal relevance:

- (1) A is causally relevant to B iff A is sufficient for B and the instances of A and B differ and are spatiotemporally proximate.

(1), as well as the other regularity theoretic accounts discussed in this chapter, heavily relies on the notions of differing instances of factors and of spatiotemporal proximity. Both of these notions call for clarifications, which, however, shall be postponed at this point. Identity criteria for instances of factors will be developed in section 2.2.2, while sections 3.2 and 5.5.2 will be concerned with spatiotemporal proximity.¹⁰ A mere pre-theoretic and intuitive understanding of when causal relata are identical as opposed to different and proximate as opposed to distant satisfactorily meets the requirements of these introductory considerations.

The simple identification of type level causes and sufficient conditions that are instantiated nearby their conditioned factors as exemplified in (1) is the target of many well known criticisms of regularity accounts.

1.2.2 *Imperfect Regularities*

There is an obvious first objection to this overly simple variety of a regularity analysis: Most causes plainly are not sufficient for their effects in the sense of (1) or, as Hitchcock (2002) puts it, causal regularities commonly are *imperfect regularities*. It is even highly dubious whether there exist any universal regularities as required by (1) at all. Consider, for instance, the striking of a match as a cause of the corresponding match catching fire. There is no doubt that striking matches indeed causes these matches to light. However, by no means every struck match actually lights. Wet matches or matches not exposed to adequate amounts of oxygen do not light even if struck repeatedly. There are causes that are not sufficient for their effects. A factor A being a sufficient condition of a factor B such that the instances of

⁸ Hume's "followed" is here – in accordance with the usual practice – understood in terms of spatiotemporal proximity such as not to preclude the possibility of simultaneous causation on a priori grounds (cf. p. 92).

⁹ For a definition of the notion of a sufficient condition see chapter 3, p. 88.

¹⁰ Prima facie, spatiotemporal proximity of causes and effects might be spelled out in terms of spatiotemporal contiguity. Yet, while contiguity certainly is sufficient for proximity it will turn out not to be necessary thereof. Depending on the level of specification, events may well be causally dependent without being spatiotemporally contiguous. Note that the notion of spatiotemporal proximity does not covertly introduce a causal concept into the analysis of (1). Proximity is a symmetric relation, whereas causal relevance is not.

A and B are spatiotemporally proximate, thus, clearly is not a necessary condition of A being causally relevant to B .

Hume did not ignore the fact that factors whose instances are not universally correlated may nonetheless be dependent in terms of general causation. In this respect, (1) does not fully reproduce Hume's analysis. In order for a factor A to be identifiable as a cause of a factor B , Hume did not require A to be sufficient for B simpliciter, i.e. sufficient relative to any causal background. Rather, he devised a cause to be sufficient for its effect only when "plac'd in like circumstances".¹¹ Hence, common analyses of causation in terms of sufficient conditions are supplemented by a *ceteris paribus* clause such that causes are merely required to be *ceteris paribus* sufficient for their effects. These considerations induce a modification of (1) to:

- (2) A is causally relevant to B iff A is *ceteris paribus* sufficient for B and the instances of A and B differ and are spatiotemporally proximate.

However, the notion of a *ceteris paribus* clause is notoriously vague. (2) is only fruitfully applicable given a proper explication of the *ceteris paribus* proviso. Chapter 5 will be concerned with spelling out what background conditions are required to coincide such that speaking of "like circumstances" is warranted. For now, a rough idea of what traditionally is meant by a *ceteris paribus* clause will suffice. In order to provide such a rough idea, consider a match that is struck against a matchbox and that, as a consequence thereof, catches fire. Refer to this scenario as " S_1 ". What requirements does a second scenario S_2 have to satisfy such that Hume would classify S_1 and S_2 as "like circumstances"? Obviously, no two scenarios coincide with respect to all their properties or characteristics. At least in its spatiotemporal properties any scenario (or circumstance) S_2 diverges from S_1 . For Hume to speak of "like circumstances" S_1 and S_2 only have to share certain significant properties. The match in S_1 is struck with a certain speed and thrust, it is exposed to a certain amount of friction, and its flammable head is dry. Moreover, S_1 features the presence of enough oxygen. If S_2 coincides with S_1 relative to these kinds of properties, only a proper subset of which have been explicitly included in the above list, S_2 can be said to satisfy the *ceteris paribus* clause with respect to S_1 , i.e. S_1 and S_2 can be referred to as "like circumstances". The properties that have to coincide in like circumstances share a common feature: They are all causally relevant to the effect under consideration. Amending Hume's definition of the notion of a cause by a thus understood *ceteris paribus* clause yields: "We may define a cause to be an object, followed by another in a first scenario S_1 , and where all the objects, similar to the first, are followed by objects similar to the second in circumstances that agree with S_1 as regards all causally relevant features". Correspondingly, (1) and (2) would have to be modified to:

- (3) A is causally relevant B iff there is a scenario S_1 such that A is sufficient for B in S_1 as well as in all scenarios that agree with S_1 as regards all causally

¹¹ Cf. Hume (1978 (1740)), p. 105.

relevant features and the instances of A and B differ and are spatiotemporally proximate.

Explicating the *ceteris paribus* proviso in this vein, of course, gives immediate rise to circularity objections.¹² (3) cannot be considered an analysis of the basal causal notion any longer, for the definiens itself presupposes the notion of causal relevance. In order to determine relative to which features of background conditions a factor A is sufficient for a factor B , (3) calls for clarity on the causes of B , which is just what (3) pretends to provide at the same time. Hence, integrating an explication of the *ceteris paribus* proviso along the lines of (3) into an analysis of causal relevance is not feasible. Nonetheless, the above considerations reveal an important feature of causal dependencies: They are not one-to-one, but many-to-one dependencies. Or put differently, while effects correspond to single factors, causes are complexes of *jointly* instantiated factors. Consequently, striking a match is not autonomously sufficient for the match to catch fire. Rather, factors as striking a match with a certain speed and thrust, dryness of its flammable head, presence of enough oxygen etc. are *jointly* sufficient for the match to light. Among the instances of these factors very specific spatiotemporal relations must subsist in order for their combination to actually become causally sufficient. Yet, postponing this problem for now,¹³ the *ceteris paribus* proviso can now be suitably accounted for without explicitly having to integrate it into an analysis of causal relevance. For if a cause is no longer held to be autonomously sufficient for its effect, but is taken to be *a mere part* of a sufficient condition, the *ceteris paribus* clause can be dropped from (2) and (3).

- (4) A is causally relevant to B iff A is a part of a sufficient condition of B and the instances of A and B differ and are spatiotemporally proximate.

As long as sufficient conditions are simply understood to be antecedents of conditionals of the form $\phi \rightarrow \psi$, the notion of *a part* of a sufficient condition is straightforwardly explicable in terms of *conjunctions* of such antecedents.¹⁴ (4) is not refuted by a struck match that does not catch fire. Whenever a match is struck, but fails to light, it may now be argued that – notwithstanding the striking – not all factors of the corresponding complex sufficient condition for lighting matches have been instantiated on the respective occasion.

1.2.3 Monotony

Implementing regularities along the lines of (4) to identify generic causal dependencies still does not amount to a feasible analysis of causal relevance, because there are regularities of the required type that are not amenable to a causal interpretation. One such type of regularities is due to the law of monotony: Antecedents

¹² Cf. Brand and Swain (1970), p. 222.

¹³ Cf. sections 2.3.4 and 3.3.

¹⁴ This simple analysis of sufficient conditions in terms of mere propositional conditionals will prove to be inappropriate in section 3.2, but for now it suffices.

of conditionals can *salva veritate* be conjunctively supplemented by further factors. Monotony allows for arbitrarily constructing complex sufficient conditions that are by no means causally interpretable. Consider again the match example discussed above. Striking a match with a certain speed and thrust, factor A , dryness of its flammable head (B), and presence of enough oxygen (C), shall be assumed to be jointly sufficient for the corresponding match to catch fire (D), i.e. $A \wedge B \wedge C \rightarrow D$. Yet, if A , B , and C are jointly sufficient for the match to light, the combination of $A \wedge B \wedge C$ and singing a song is thus sufficient, too. Moreover, $A \wedge B \wedge C$ combined with singing a song and blinking an eye and wiggling the left pinky toe are also going to be jointly sufficient for the match to catch fire. Or formally:

$$A \wedge B \wedge C \rightarrow D \vdash A \wedge B \wedge C \wedge X \rightarrow D, \quad (1.1)$$

where X stands for an arbitrary factor or conjunction of factors.¹⁵ This demonstrates that being a part of a sufficient condition, i.e. being a conjunct within a sufficient conjunction of factors, is by no means sufficient for being causally relevant to the respective conditioned factor.

Broad (1930) has been the first to propose a solution to this problem. He does not analyze causes to be mere parts of sufficient conditions, but, rather, to be *non-redundant* parts of such conditions. A non-redundant part of a sufficient condition can be spelled out – in purely logical terms¹⁶ – as being a conjunct of a sufficient condition such that, if it is eliminated from that condition, the latter loses its sufficiency for the corresponding effect. Complex causes then are no longer understood as conjunctions of factors which are jointly merely sufficient for their effect, but are newly taken to be *minimally sufficient* conjunctions of their effects – a minimally sufficient conjunction being a conjunction that does not have sufficient proper parts or, in other words, that cannot be reduced by any conjunct without loss of sufficiency.¹⁷

- (5) A is causally relevant to B iff A is a part of a minimally sufficient condition of B and the instances of A and B differ and are spatiotemporally proximate.

Applying (5) to the match example prohibits a causal interpretation of, say, the combination of striking a match, presence of enough oxygen, dryness of the match, and singing a song. The conjunction of these factors is merely sufficient, but not minimally sufficient for the match to catch fire. One of its conjuncts, the singing, can be eliminated without loss of sufficiency. Requiring a minimalization of sufficient conditions in terms of Broad precludes a causal interpretation of arbitrary extensions of sufficient conditions based on the law of monotony.

¹⁵ X , X_1 , X_2 etc. are variables running over conjunctions of factors (cf. section 3.5.)

¹⁶ Not all critics of regularity accounts have taken note of these purely logical ways to minimalize sufficient conditions. For instance, in 1970 Brand and Swain still erroneously claimed that minimalizing sufficient conditions cannot be accomplished in non-causal (and thus non-circular) terms (cf. Brand and Swain (1970), p. 226).

¹⁷ For details on the notion of minimal sufficiency see section 3.3.

1.2.4 Empty Regularities

Minimalizing sufficient conditions does not solve all the problems that can be induced by the law of monotony. Consider again the match example. As we have seen, the presence of oxygen, factor C , is not itself sufficient for a match to catch fire (D). Other factors – A and B – have to be instantiated as well in order for an instance of D to occur. Yet, $A \wedge B \wedge C$ is not the only minimally sufficient condition containing C . Another such condition is constituted by the presence of oxygen and the absence of oxygen: $C \wedge \neg C$. A contradiction is sufficient for any factor, not only for matches catching fire, but also for rain to fall and elephants to be born. That $C \wedge \neg C$ is moreover minimally sufficient for a match to light can easily be verified by either removing C or $\neg C$, both of which is accompanied by a loss of sufficiency for D . More generally put: Material conditionals are true if their antecedents are false or non-instantiated, or *empty* for short. Any regularity statement as resorted to in (5) whose antecedent is empty is, accordingly, termed an *empty regularity*. Empty regularities do not only result from contradictory antecedents and, thus, from logically non-instantiatable antecedents, but also from physically non-instantiatable antecedents as, for instance, “Whenever Pegasus goes skiing, Lake Thun is made of gold”.

The truth of empty regularity statements raises another often cited problem for regularity accounts: Empty regularities are, notwithstanding their truth, not amenable to a causal interpretation.¹⁸ The combination of absence and presence of oxygen – $C \wedge \neg C$ – is no causally relevant to the sinking of Mississippi steamers, even though $C \wedge \neg C$ in fact is minimally sufficient for these sinkings. Neither can Pegasus’ ski tour be seen as a cause of the golden content of a lake.

Solutions to this problem are easily thought of. It is not the case that only a certain proper subset of all empty regularities consists of regularities that are not causally interpretable, rather, *no* empty regularities are thus interpretable. Causal dependencies only subsist among causes and effect that are instantiated in nature.¹⁹ “Instantiated” in this context is not to be read in terms of “has occurred prior to a specific moment of investigation”, but rather in terms of “has not occurred in all past and will not occur in all future”. Relative to a Humean analysis, factors and conjunctions of factors may not be causally related if they have not been instantiated in all past and will not be instantiated in all future. Therefore, empty regularities can straightforwardly be excluded from causal interpretability by adding a further constraint to (5) that requires the antecedent of causally interpretable regularities to be non-empty.

(6) A is causally relevant to B iff the following conditions hold:

- (i) A is a part of a minimally sufficient condition X_1 of B ,

¹⁸ Cf. e.g. Kneale (1961), Armstrong (1983), chs. 2, 5, 8, van Fraassen (1989b), ch. 5.

¹⁹ Note moreover that regularity theories of *natural laws* are not as easily rendered immune to the empty regularities problem as regularity theories of causation. While factors without instances may not be claimed to be causally dependent, there may well be natural laws involving predicates with empty extensions, as e.g. “... travels faster than light” (cf. Molnar (1969)).

- (ii) the instances of A and B differ and are spatiotemporally proximate, and
- (iii) there is an instance of X_1 .

Alternatively it might be argued that event types or conjunctions of event types without instances must not even be taken into consideration in the first place when it comes to causal analyses. Thus, existence requirements with respect to causally analyzed factors might be imposed as a kind of criterion of well-formedness for causal factors and conjunctions thereof. Upon opting for this solution to the empty regularities problem, which essentially amounts to the same as a solution in the vein of (6), an explicit modification of (5) can even be dispensed with by simply relativizing regularities as used in (5) to well-formed causal factors.

1.2.5 Non-Symmetry

Another objection often raised against regularity accounts concerns the direction causation. Causation is not symmetric and this lacking symmetry is usually claimed not to be adequately representable in regularity theoretic terms. In order to clarify what this criticism amounts to, some conceptual preliminaries are required. First, *non-symmetry* must be distinguished from *asymmetry*. A relation C is non-symmetric iff (1.2) holds and asymmetric iff (1.3) holds.²⁰

$$\neg\forall x\forall y(Cxy \rightarrow Cyx) \text{ or equivalently } \exists x\exists y(Cxy \wedge \neg Cyx) \quad (1.2)$$

$$\forall x\forall y(Cxy \rightarrow \neg Cyx) \quad (1.3)$$

Every asymmetric relation with a non-empty extension is also non-symmetric, but not vice versa. Predicating asymmetry of a non-empty relation is a much stronger claim than predicating non-symmetry of it. Second, questions as to the symmetry of *general* and *singular causation* must be clearly distinguished, as type and token level causation do not share all relational properties. Some such properties are beyond doubt, others are controversial.

Both general and singular causation are non-symmetric relations. There exist both general and singular causal dependencies and neither “Drinking is a cause of drunkenness” implies “Drunkenness is a cause of drinking” nor “Tom’s drinking of 6 beers at noon on September 7, 2004, causes Tom’s drunkenness in the afternoon of September 7, 2004” implies “Tom’s drunkenness in the afternoon of September 7, 2004, causes Tom’s drinking of 6 beers at noon on September 7, 2004”. Moreover, general causation is not asymmetric, for there are event types that are causally relevant to themselves, as e.g. in causal cycles. For instance, with increasing unemployment the consumption of the population is reduced. This causes decreased profits on the side of the employers, which, in turn, causes them to lay off even

²⁰ This terminology corresponds to logical and mathematical conventions (cf. e.g. Lemmon (1978 (1965)), pp. 180-182). In addition to non-symmetry and asymmetry a relation may lack symmetry in terms of so-called *antisymmetry*. A relation C is antisymmetric iff $\forall x\forall y(x \neq y \wedge Cxy \rightarrow \neg Cyx)$.

more people. Thus, the unemployment increases anew. In contrast, on token level there are no events that cause themselves. Contrary to general causation, singular causation is irreflexiv. Yet, whether it is moreover asymmetric is a question that is heavily disputed in the literature. If event a is a cause of event b , does it follow that b is not a cause of a ? Is there reciprocal causation on token level?²¹ We can postpone this question until section 2.3.1, for a regularity theory that focuses on general causation primarily needs to account for the latter's non-symmetry. In consequence, the criticism against regularity theories to the effect that they are not capable of adequately capturing the direction of causation amounts to the claim that the non-symmetry of general causation cannot be mirrored on regularity theoretic grounds. And indeed, *prima facie* there appears to be a problem.

A being causally relevant to B neither implies B being causally relevant to A nor \bar{B} being causally relevant to \bar{A} . However, by contraposition $A \rightarrow B$ is a true regularity statement iff $\neg B \rightarrow \neg A$ is so too. Which of these conditionals is to be causally interpreted? It is certainly not the case that a factor is causally relevant to another factor iff the negation of the latter is causally relevant to the negation of the former. Accordingly, many critics of regularity theories have claimed that regularity accounts cannot adequately distinguish between causes and effects.²² (6) identifies A to be causally relevant to B iff it identifies \bar{B} to be causally relevant to \bar{A} , which indicates that we have not come up with an adequate analysis of causal relevance yet.

Satisfactorily mirroring the non-symmetry of general causation is an intricate problem not only faced by regularity accounts, but by virtually all presently known theories of causation. Normally the direction of causal relevance is accounted for with recourse to some non-symmetry *external* to the conceptual framework used in the analyses of causal relevance as – most prominently – the direction of time or human manipulation and intervention.²³ Applied to the regularity theory considered here, this could possibly yield that conditionals in the sense of (6) are causally interpretable only if the instances of the antecedent *precede* the instances of the consequent. Along these lines, one of $A \rightarrow B$ and $\neg B \rightarrow \neg A$ could be excluded from causal interpretability. However, in section 3.6.2 we shall see that accounting for the non-symmetry of general causation by means of an external non-symmetry is a very high theoretical price to pay and, moreover, is not necessary in order to represent the direction of causation.

General causation can be oriented on mere logical grounds. Roughly, while conditional dependencies among single factors cannot be attributed a direction without resorting to external non-symmetries, complex *nets* of such dependencies can be oriented based on existing regularities only. There are several alternative type level causes for each effect. A match can be lit by either striking it

²¹ Cf. Eells (1991), ch. 5, Hausman (1998), pp. 44-47.

²² Cf. e.g. Armstrong (1983), ch. 2.

²³ Cf. Hume (1999 (1748)), p. 146, Suppes (1970), p. 23, Gasking (1955), von Wright (1971), Price (1992).

against a match box, by exposing it to fire or to a flammable chemical etc. Accordingly, causally interpretable regularities are far more complex than expressed by (6). Rather than merely one minimally sufficient condition $A \wedge C \wedge D$, many alternative minimally sufficient conditions – $A \wedge C \wedge D$, $E \wedge F \wedge G$, $H \wedge I \wedge J$, ... – must be invoked for each effect. Moreover, an effect does not occur without the presence of at least one of its alternative causes. Thus, whenever the effect is given, at least one of its alternative minimally sufficient conditions is instantiated as well. These mutual dependencies among causes and effects are tentatively²⁴ expressible by means of a biconditional as in (1.4).

$$(A \wedge C \wedge D) \vee (E \wedge F \wedge G) \vee (H \wedge I \wedge J) \leftrightarrow B \quad (1.4)$$

Each complex cause of B is minimally sufficient for B , while the disjunction of all alternative causes is *necessary* for B .²⁵ (1.4) is non-symmetric with respect to the expressions to the left and the right of “ \leftrightarrow ”. The instantiation of a particular disjunct is minimally sufficient for B , but not vice versa. B does not determine a particular disjunct to be instantiated.²⁶ B only determines the whole disjunction of minimally sufficient conditions. $A \wedge C \wedge D$ and $E \wedge F \wedge G$ and $H \wedge I \wedge J$ are each minimally sufficient for B , the latter however is only minimally sufficient for $(A \wedge C \wedge D) \vee (E \wedge F \wedge G) \vee (H \wedge I \wedge J)$. Hence, given that an instantiation of $A \wedge C \wedge D$ is observed, it can be inferred that there is an instance of B somewhere in the corresponding spatiotemporal neighborhood. On the other hand, if an instance of B is observed, no such inference to a proximate instantiation of $A \wedge C \wedge D$ is possible. The observed instance of B might well have been caused by $E \wedge F \wedge G$. This non-symmetry corresponds to the non-symmetry of determination. It induces a specification of (6) along the following lines:

- (7) A is causally relevant to B iff the following conditions hold:
- (i) A is a part of a minimally sufficient condition X_1 of B ,
 - (ii) X_1 is a disjunct contained in a disjunction $X_1 \vee X_2 \vee \dots \vee X_n$, $n \geq 2$, of other minimally sufficient conditions of B , such that $X_1 \vee X_2 \vee \dots \vee X_n$ is necessary for B ,
 - (iii) the instances of X_1 and B differ and are spatiotemporally proximate, and
 - (iv) there is an instance of X_1 , X_2 , ..., and of X_n .

²⁴ (1.4) is a mere tentative formal representation because propositional logic does not allow for adequately expressing the relational constraints implicit in causal regularities in the sense of (6). Chapter 3.5 will therefore resort to first-order logic in order to adequately account for causal regularities on logical grounds.

²⁵ This essentially corresponds to Mackie’s (1974) famous analysis of causation in terms of so-called *INUS-conditions*. Mackie (1974) will not be given an in-depth review in the present context. This has been done in Baumgartner and Graßhoff (2004), ch. 5.

²⁶ Cf. Graßhoff and May (2001), pp. 97-99. Similar analyses of the direction of causation have been proposed in Russell (1992 (1913)), p. 199, Sanford (1976), Ehring (1982), and Hausman (1998).

Clearly though, by contraposition (1.4) is equivalent to

$$\neg B \leftrightarrow \neg(A \wedge C \wedge D) \wedge \neg(E \wedge F \wedge G) \wedge \neg(H \wedge I \wedge J) \quad (1.5)$$

However, in view of the fact that effects have several alternative causes, (7) restricts the causal interpretability of complex regularity statements to one specific syntactical form. Within a set of logically equivalent regularity statements, only expressions with a syntax that exhibits alternative minimally sufficient conditions as *disjuncts* of a necessary condition are causally interpretable. Applied to (1.4) and (1.5), this syntactical constraint prohibits a causal interpretation of (1.5), for it does not render an underlying causal structure transparent in the sense just delineated.²⁷

(1.5) is moreover equivalent to a biconditional that results from (1.5) by factoring out and bringing the righthand side back into disjunctive normal form:

$$(\neg A \wedge \neg E \wedge \neg H) \vee (\neg A \wedge \neg E \wedge \neg I) \vee \dots \vee (\neg D \wedge \neg G \wedge \neg J) \leftrightarrow \neg B \quad (1.6)$$

In contrast to (1.5), (1.6) is unproblematically causally interpretable. While (1.4) identifies three minimally sufficient conditions as complex causes of B , (1.6) establishes the causally interpretable minimally sufficient conditions of $\neg B$. Each of those conditions amounts to a conjunction consisting of the negation of exactly one conjunct of each disjunct of (1.4). Furthermore, (1.6) does not reverse the direction of the causal dependencies expressed in (1.4). Both identify B and $\neg B$, respectively, as effects and the other factors as causes. Thus, there is one regularity statement complying to the syntactical constraints imposed by (7) for a positive effect and one for the latter's negative complement. Both of these regularities exhibit the same non-symmetry. Accordingly, neither of them poses a problem for (7).

Accounting for the non-symmetry of general causation in this vein has an important implication as regards the minimal complexity of causal structures. A factor or conjunction of factors X_1 , that is both minimally sufficient and necessary for another factor or conjunction of factors X_2 , cannot be identified as cause of X_2 , for X_2 would be minimally sufficient and necessary for X_1 as well. All empirical evidence such a dependency structure would generate are perfectly correlated instantiations of X_1 and X_2 – both would either be co-instantiated or absent. Such empirical data is not causally interpretable. In order to distinguish causes from effects and to orient the relation of causal relevance, at least *two* alternative causes are needed for each effect.

This rough sketch of the analysis of the direction of general causation to be proposed in the study at hand will be further elaborated in section 3.6.2. For now, it suffices to note that, contrary to the widespread opinion in the literature, regularity

²⁷ The fact that logically equivalent expressions differ with respect to the straightforwardness of their causal interpretation is analogous to the fact that divergent normal forms differ with respect to how transparent they render truth-conditions of logical formulas. For instance, it is much more intricate to read off truth-conditions from prenex normal forms than, say, from normal forms with minimal quantifier scopes (cf. Hintikka (1973)).

theories not only adequately account for the non-symmetry of general causation, but moreover succeed in doing so without resorting to non-symmetries external to the conceptual framework implemented in their analyses of causal relevance. Against this background, such external non-symmetries as the direction of time or of human intervention remain amenable to a straightforward analysis in terms of the non-symmetry of general causation. Section 3.6.2 will further substantiate that mirroring the non-symmetry of general causation is not a problem, but, rather, one of the most salient qualities of regularity accounts.

1.2.6 Spurious Regularities

One of the most widespread criticisms against regularity theories stems from so-called *spurious regularities*.²⁸ Consider two parallel effects A and B of a common cause C and assume for simplicity's sake that C in fact is minimally sufficient for A and B .²⁹ Such as to do justice to the complexity of causal structures let us suppose there exists one minimally sufficient alternative cause for A and B each – D for A and E for B . All in all, the causal structure under consideration thus is assumed to be of a form as depicted in figure 1.1.³⁰ In this constellation, A in combination with the absence of D , i.e. $A \wedge \neg D$, is minimally sufficient for B without $A \wedge \neg D$ being a complex cause of B . Whenever $A \wedge \neg D$ occurs, C is present as well, for no effect occurs without any of its causes. Hence, if D is absent, C must be present to account for A . Furthermore, since C is taken to be sufficient for B , it follows that $A \wedge \neg D$ is sufficient for B as well. Of course, $A \wedge \neg D$ is moreover part of a necessary condition of B :

$$(A \wedge \neg D) \vee C \vee E \leftrightarrow B \quad (1.7)$$

According to (7), (1.7) is a regularity statement that is causally interpretable. This clearly is an unacceptable consequence, for, as mentioned above, relative to the construction of the structure in figure 1.1, $A \wedge \neg D$ is not causally relevant to B .

Structures as the one under consideration are ubiquitous. The most famous concrete example of this type is the so-called *Manchester-Factory-Hooters* example based on which Mackie (1974) ultimately abandoned the attempt to provide a

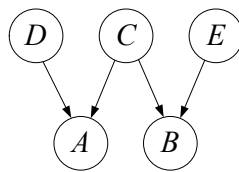


Fig. 1.1: An common cause structure that gives rise to spurious regularities.

²⁸ Cf. e.g. Cartwright (1989), pp. 25-29.

²⁹ For details on common cause structures see section 2.3.

³⁰ The graphical notation implemented here is properly introduced in section 2.3.2.

genuine regularity theoretic analysis of causation.³¹ Examples of this type unmistakably demonstrate that necessary conditions, just as sufficient conditions, may contain redundancies. A being necessary for B implies that $A \vee C$ is necessary for B . Or formally:

$$B \rightarrow A \vdash B \rightarrow A \vee C. \quad (1.8)$$

Any true conditional stays true if any (true or false) disjunct is added to its consequent. Analogous to the case of sufficient conditions, the extendability of necessary conditions by arbitrary disjuncts forecloses a causal interpretability of necessary conditions. A causal interpretation of necessary conditions is only warranted if the conditions exclusively contain factors that are essential to the bringing about of the purported effect. Arbitrary factors as C in (1.8) or conditions as $A \wedge \neg D$ in (1.7) must, even if they are minimally sufficient, not be incorporated in causally interpretable necessary conditions.

A solution to this problem that is analogous to the solution of the difficulties induced by monotony as been proposed in May (1999) and Graßhoff and May (2001): Necessary conditions must be minimalized. The basic idea behind the minimalization of necessary conditions coincides with the criterion guiding the minimalization of sufficient conditions: A necessary condition is *minimally* necessary iff it does not contain a necessary proper part. Minimalizing necessary conditions based on this notion of a minimally necessary condition in fact eliminates all and just the spurious minimally sufficient conditions as $A \wedge \neg D$ from complex regularity statements as (1.7). In order to see this, consider again the structure depicted in figure 1.1: Whenever B is given, C or E is instantiated. Thus, $C \vee E$ is necessary for B . The antecedent of (1.7) has no other necessary proper part. $(A \wedge \neg D) \vee C$ is not necessary, for there are instances of B without $A \wedge \neg D$ and C being instantiated – say, when $A \wedge D$ is given along with $\neg C$ and E . Neither is $(A \wedge \neg D) \vee E$ necessary for B : There are instances of B without instances of $A \wedge \neg D$ and E occurring – for example, when $A \wedge D$ is given in combination with $\neg E$ and C . Among the elements of the necessary condition of B mentioned in (1.7) the following asymmetry holds, which allows for eliminating $A \wedge \neg D$: $A \wedge \neg D$ implies $C \vee E$, while $C \vee E$ does not imply $A \wedge \neg D$. That means, $C \vee E$ is a minimally necessary disjunction of minimally sufficient conditions of B , or formally:

$$(C \vee E \rightarrow B) \wedge (B \rightarrow C \vee E) \wedge \neg(B \rightarrow C) \wedge \neg(B \rightarrow E). \quad (1.9)$$

Thus, there in fact exists a suitable refinement of (7) that allows for an appropriate handling of spurious regularities on regularity theoretic grounds.

(8) A is causally relevant to B iff the following conditions hold:

- (i) A is a part of a minimally sufficient condition X_1 of B ,

³¹ Cf. Mackie (1974), p. 83 et seq., Cartwright (1989), pp. 25-29. In Baumgartner and Graßhoff (2004), pp. 99-103, we have discussed the Manchester-Hooters in all detail along with a solution to this problem.

- (ii) X_1 is a disjunct contained in a disjunction $X_1 \vee X_2 \vee \dots \vee X_n$, $n \geq 2$, of other minimally sufficient conditions of B , such that $X_1 \vee X_2 \vee \dots \vee X_n$ is *minimally* necessary for B ,
- (iii) the instances of A and B differ and are spatiotemporally proximate, and
- (iv) there is an instance of X_1 , X_2 , \dots , and of X_n .³²

1.2.7 Single-Case Regularities

A problem that is closely related to the problem of empty regularities has e.g. been raised by Armstrong (1983).³³ A conditional turns out true if both its antecedent and consequent are true. Thus, if antecedent and consequent of a conditional each report the occurrence of a singular event that actually has occurred, the conditional as a whole is true. Therefore, Armstrong argues, a regularity as required by a regularity theory subsists among any two factors with a single instance each, irrespective of whether they are causally related or not. No doubt, a conditional as “Whenever Nero sets fire on Rome, the Titanic sinks” is true and, no doubt, we are not prepared to hold Nero causally responsible for the sinking of the Titanic. Hence, Armstrong’s argument continues, not only empty, but also these so-called *single-case* regularities pose a serious problem for a regularity theory.

At first, it must be pointed out that the plain truth of a conditional as “Whenever Nero sets fire on Rome, the Titanic sinks” does not suffice to identify Nero’s setting fire on Rome as a cause of the sinking of the Titanic according to any of the regularity theoretic accounts (1) to (8) considered thus far. For these accounts not only require causes and effects to satisfy a material conditional, but moreover to be proximately instantiated. Even though the notion of spatiotemporal proximity has not been properly explicated as yet, relative to any pre-theoretic understanding of that notion, it seems plain that Nero’s setting fire on Rome and the sinking of the Titanic cannot be seen as proximate events. However, this shortcoming of Armstrong’s argument is easily remedied. Assume that Harold Bride, the junior wireless operator on the Titanic, for the first (and only) time in his life lit a Havana cigar moments before the ship hit the iceberg. The conditional “Whenever Harold Bride lights a Havana, the Titanic sinks” is true and, moreover, the events mentioned in its antecedent and consequent are spatiotemporally proximate. Of course, Bride’s lighting of a Havana is not only sufficient, but moreover minimally sufficient for the Titanic to sink. The antecedent of the aforementioned conditional does not comprise a sufficient proper part, i.e. any removal of a factor results in the loss of sufficiency.³⁴ Furthermore, Bride’s lighting of a Havana is not the only

³² In fact, in section 3.4 we shall see that the notion of a minimally necessary condition renders condition (iv) of (8) redundant, because the addition of this notion to condition (ii) implies that each disjunct in $X_1 \vee X_2 \vee \dots \vee X_n$ is instantiated at least once.

³³ Cf. Armstrong (1983), pp. 15-17, similarly Mellor (1995), p. 15.

³⁴ Sufficient antecedents consisting of one factor only are automatically minimally sufficient (cf. section 3.3).

minimally sufficient condition for the sinking of the Titanic. The latter's real cause constituted by the collision with the iceberg amounts to another such condition. Hence, there is a necessary condition of the sinking of the Titanic that contains Bride's lighting of a Havana as a minimally sufficient disjunct. This suffices to refine Armstrong's argument such that it does justice to the complexity of causal structures as required by (7): Any two factors with exactly one instance each, such that these instances are spatiotemporally proximate, satisfy a regularity as required by (7), yet by no means all thus related factors are related in terms of causal relevance as well. Consequently, (7) does not amount to a sufficient condition for causal relevance.

That (7) is unsuited as analyses of general causation has already been demonstrated by the problem of spurious regularities. In order for Armstrong to succeed in establishing that single-case regularities prove the fundamental defectiveness of regularity accounts, his argument must be tailored to be directed against (8). It thus must be shown that Bride's lighting of a Havana is not only contained in a necessary condition of the sinking of the Titanic, but is moreover a non-redundant part of a causally interpretable minimally necessary condition thereof.

Before this further refinement will be attempted a possible objection against Armstrong's argument has to be considered. Antecedent and consequent of "Whenever Harold Bride lights a Havana cigar, the Titanic sinks" involve proper names or, if formal explications by means of definite descriptions are preferred, predicates that apply to single events only – more generally: *local predicates*, i.e. predicates that involve spacetime coordinates or singular terms. The admissibility of local predicates in contexts of general causation, as is well known, is commonly denied in the literature.³⁵ Causal dependencies do not exclusively subsist in local domains as the one constituted by the Titanic's maiden voyage. Harold Bride's lighting of a Havana cigar is a cause of the sinking of the Titanic if and only if lighting Havana cigars generally cause ocean liners to sink. In view of this universality of causal dependencies and, moreover, in light of the basic intuition behind regularity accounts according to which only repeated instantiations of factors allow for causal diagnoses, it is plain that proper names and local predicates must be excluded from well-formed factor definitions.³⁶ Armstrong's argument might thus be rejected for its involvement of factors that are defined by means of local predicates and that, accordingly, are not well-defined causal factors. Yet, as Armstrong points out, especially in macroscopic contexts as the one under consideration every predicate involving local constraints may well be replaceable by a co-extensional non-local predicate.³⁷ For instance, "... is Harold Bride's lighting of a Havana cigar" could be replaced by a conjunction of arbitrary non-local properties that, taken together, happen to apply to exactly one event, namely Harold Bride's lighting of a Havana cigar. Or instead of by use of a proper name, Harold Bride might be

³⁵ Cf. e.g. Popper (1994 (1934)), ch. 5, Hempel (1977 (1965)), ch. 2, Goodman (1983). More will be said about local predicates in section 2.2.2.

³⁶ Cf. section 2.2.2.

³⁷ Cf. Armstrong (1983), ch. 1.

referred to by specifying his genome, while the Titanic is individuatable by means of its molecular structure. So let us grant that “Whenever Harold Bride lights a Havana cigar, the Titanic sinks” represents the exact same single-case regularity as “Whenever a person with genome s lights a Havana cigar, an ocean liner with molecular structure t sinks”, or formally $S \rightarrow T$, which constitutes a single-case regularity involving non-local predicates and thus well-defined causal factors only.

The Titanic’s collision with the iceberg, of course, is expressible by means of non-local predicates as well. Symbolizing this (non-locally defined) collision by C we get the following true biconditional that is causally interpretable according to (7):

$$S \vee C \leftrightarrow T. \quad (1.10)$$

If $S \vee C$ is not merely necessary, but moreover *minimally* necessary for T , (1.10) not only refutes (7), but also (8). $S \vee C$, however, does not amount to a minimally necessary condition of T . There is only one single instance of each factor involved in (1.10). Whenever T is instantiated, both S and C are instantiated nearby. Thus, the antecedent of (1.10) can be further minimalized:

$$S \leftrightarrow T \quad (1.11)$$

$$C \leftrightarrow T \quad (1.12)$$

Neither (1.11) nor (1.12), however, are causally interpretable, for these expressions merely report a perfect correlation of T and S and C . Any of these factors is instantiated if and only if the other two factors are instantiated as well. Such perfect correlations, as we have seen above, are not causally interpretable, for none of the involved factors is identifiable as cause or effect. Since (8) requires causally interpretable regularities to specify minimally necessary conditions of a certain minimal complexity, neither (1.11) nor (1.12) is amenable to a causal interpretation according to (8). That, however, does not mean that (8) does not allow for identifying the collision with the iceberg as a cause of the sinking of the Titanic. The impossibility to causally interpret either (1.11) or (1.12) merely indicates that the events involved in the sinking of the Titanic are typecast in an overly fine-grained manner. Causal structures can only be unfolded when a sufficient amount of comparable test situations are available. Thus, the Titanic’s sinking must be placed in a broader context, it must be typecast in a more coarse-grained fashion, e.g. as a sinking of an ocean liner. Such a typing will immediately yield far more instances for each causal factor, which, in turn, will suspend biconditional dependencies as in (1.11) or (1.12). If not only the Titanic’s sinking, but ocean liner sinkings are taken into consideration, many cigar lightings will be found that are in no way followed by a sinking ocean liner, whereas collisions with icebergs either retain their sufficiency for ocean liner sinkings or a subsequent expansion of the set of investigated factors will reveal them to be parts of minimally sufficient conditions of such sinkings.

All in all, a regularity theoretic analysis of general causation imposes certain minimal complexity constraints on causal structures. Causal structures are not

one-to-one dependencies among single factors. Every effect has several alternative complex causes. In order to establish such dependencies, more than one single instance of causes and effects is required. Chapter 5 will be concerned with the question as to how many instances of causal factors are needed to identify a causal structure of a given complexity. What matters for now is that there are regularity theoretic notions of causation – e.g. along the lines of (8) – that mirror these minimal complexity requirements and that, accordingly, are not affected by the problem of single-case regularities.

1.2.8 Indeterminism

The discussion of the thus far considered arguments against regularity accounts revealed that common objections attack overly simplistic regularity theoretic analyses. None of these objections affects a regularity account – as will be developed in the following – that adequately represents the whole complexity of causal structures. Still, there is one conventional argument often put forward against regularity theories that does not target an oversimplified account. With the advent of quantum mechanics, it appears that some processes as e.g. radioactive decay both run irreducibly indeterministically and are to be qualified as being of causal nature. Accordingly, the second half of the 20th century has seen the rise of probabilistic analyses of causation as for instance proposed by Reichenbach (1956), Good (1961), and Suppes (1970), who took these indeterministic processes as sufficient evidence to the effect that building a theory of causation on the principle of determinism³⁸, as regularity accounts generally do, is fundamentally mistaken.

While the previously considered objections do not question the principle of determinism and thereby the conceptual fundament of regularity accounts, that is just what this argument from indeterminism does. Criticizing regularity theories in this vein presupposes that there are causal processes that run irreducibly indeterministically. While the existence of irreducibly indeterministic processes is hardly challengeable according to standard interpretations of quantum mechanics³⁹, there are many open questions – as for instance raised by phenomena of the EPR type⁴⁰ – with respect to the causal interpretability of these processes. The causal interpretability of indeterministic processes is even more questionable as not even modern probabilistic theories of causation can successfully account for these processes. The latter violate central assumptions of probabilistic analyses of general causation, as the Reichenbachian common cause principle⁴¹ or the causal Markov assumption.⁴² In consequence, probabilistic accounts of general causation are forced to limit their scope to so-called pseudo-indeterministic processes,

³⁸ Cf. section 2.4.

³⁹ Nonetheless, there also are deterministic interpretations of quantum mechanics (cf. Albert (1992)).

⁴⁰ Cf. van Fraassen (1989a).

⁴¹ Cf. Reichenbach (1956).

⁴² Cf. Spirtes, Glymour, and Scheines (2000 (1993)).

i.e. processes whose indeterminacy is merely due to incomplete knowledge of or control over the involved factors.⁴³

The question as to whether the irreducibly probabilistic progression of quantum mechanical processes poses a problem to a regularity account of causation in the end boils down to what pre-theoretic understanding of the notion of a cause-effect relation is presupposed. Or put differently: Irreducibly probabilistic processes pose a problem to a regularity account only if such processes are to be identified as *causal* processes. If, however, the notion of a causal dependency is taken to be essentially tied to such principles as the principle of determinism, quantum mechanical processes are not classified as causal to begin with. Against such a conceptual background, indeterminism, rather than compromising regularity accounts, raises questions as to how it can “be the case in an indeterminist world that some events are causally determined while others are not”.⁴⁴ Indeed, in view of irreducibly indeterministic processes, the regularity theorist might well retreat to a more moderate position according to which several types of causal relations can be distinguished, one of which – prevalent in macroscopic contexts – being a deterministic relation. Then he could propagate his account as analysis of just that deterministic causal relation.⁴⁵

The study at hand will not be concerned with metaphysical or conceptual matters such as the question whether there in fact are irreducibly indeterministic processes in nature that we want to classify as causal or not. If this question is positively answered, no presently known theory of causation that is located in the empiricist Humean tradition can claim to satisfactorily handle these irreducibly indeterministic dependencies. The present inquiry will take for granted the validity of the principle of determinism and, accordingly, will only be concerned with a kind of causal relation that satisfies this principle. Whether there are other forms of causal dependencies will be left open here.

1.3 Outline

Chapter 2 provides the conceptual fundament of the subsequent regularity theory which is not affected by the aforementioned objections and which does justice to the claims put forward in section 1.1. An evaluation of the debate over fact vs. event causation will show that analyzing the relata of singular causation in terms of facts has unclear and very broad philosophical follow-ups, whereas taking causes and effects to be events offers grounds to meet all requirements of a successful analysis of the token level causal relata without such uncontrollable theoretical

⁴³ Cf. Spirtes, Glymour, and Scheines (2000 (1993)), Davis (1988) or Cartwright (1999). For further details on these matters see Baumgartner and Graßhoff (2004), ch. 6, and Manni (forthcoming).

⁴⁴ Cf. Xu (1997), p. 137. Accordingly, Xu (1997) proposes a deterministic analysis of causation that he claims to be compatible with indeterminism. Similarly: Belnap (2005).

⁴⁵ Confining oneself to such a moderate position seems advisable for determinists in view of arguments in favor of a causal interpretation of irreducibly indeterministic processes as presented in Lewis (1986), p. 176, or Mellor (1995), ch. 5.

spin-offs. Furthermore, chapter 2 introduces the notion of *causal relevance* – the central analysandum of a regularity theory –, reviews the atomic forms of generic causal dependencies, and discusses the causal principles presupposed by a regularity account.

The core of chapter 3 consists in the analysis of causal relevance, i.e. of general causation. We shall see that (8) is still defective in several respects. As indicated above, causal regularities will be accounted for by means of first-order logic and the notion of a *minimal theory* will be introduced as a conceptual skeleton of our analysis. Apart from clarifying causal relevance subsisting among positive factors, chapter 3 will account for causation among absences or negative factors, for the direction of causation, for variable specifiability of causal structures, and for the construction of complex structures from a limited set of atomic causal dependencies. By drawing on this notion of general causation, chapter 3 will, moreover, spell out singular causation.

In chapter 4 I then call attention to a problem arising from ambiguities as regards the composition of complex structures, especially causal chains, from simple ones. Up to recent years causal chains were given surprisingly little attention in studies on causation. Philosophical analyses of causation focussed their interest on direct dependencies among causes and effects. Apparently, for a long time it was generally assumed that complex structures could be straightforwardly accounted for once a successful analysis of atomic causal dependencies would be available, i.e. of dependencies among single factors. However, as I shall show in chapter 4, this confidence has been premature for many prominent theoretical accounts of causation, including regularity theories. It will be demonstrated that for every chain there exists an empirically equivalent common cause structure. That means, there is no empirical warrant for the existence of one of the most central and ubiquitous causal structures: causal chains! Chapter 4 will then propose a solution to this problem and provide the theoretical basis for a procedure that maps complex causal structures onto sets of coincidentally instantiated factors.

Finally, chapter 5 develops a procedure of causal reasoning based on the results obtained in chapter 4. An algorithm of causal reasoning is proposed that assigns complex structures to sets of coincidentally instantiated factors. Contrary to existing procedures, this so-called *coincidence analysis* does not build up complex structures layer by layer by suitably combining atomic causal dependencies among single factors. Rather, it directly maps complex structures onto coincidence sets. Thereby the ambiguities encountered upon distinguishing between chains and common causes structures can be avoided. Moreover, coincidence analysis fills a gap left open by the probabilistic algorithms of causal reasoning as presented in Spirtes, Glymour, and Scheines (2000 (1993)). These algorithms only generate informative outputs provided that analyzed conditional probabilities are lower than 1, i.e. provided that causes do not in a strict sense determine their effects. In contrast, coincidence analysis is custom-built to deterministic causal dependencies and properly uncovers such dependencies.

2. CONCEPTUAL FUNDAMENT

2.1 Introduction

This chapter provides the conceptual fundament of the theory of causation to be developed in chapter 3. The previous chapter has shown, that – at least – two causal relations must be kept apart – one on token and one on type level. Both causal relations have their specific relata: Token events in case of singular causation, factors in case of general causation. While regularity theories take general causation to be conceptually primary, the notion of a token event is more basic than the notion of a factor, for the latter shall be understood in terms of similarity sets of the former. This chapter, thus, first accounts for the relata of singular causation and then for the relata of general causation. Moreover, it discusses the relational properties of the two causal relations and introduces four principles of causation.

In order to avoid unnecessary terminological complications, we shall often simply speak of *causation* or *causes* and *effects* whenever the context clarifies whether singular or general causation is under consideration or whenever both causal relations are at issue.

2.2 The Relata of Causation

2.2.1 Fact vs. Event Causation

The question as to what entities operate as relata of the singular causation has been controversially debated in the literature concerned with causality of the past century. Many different answers have been offered, some of which are incompatible, others only terminologically differing. The by far most popular view holds token level causes and effects to be *events*.¹ In opposition to the event camp, a number of philosophers, among which Horgan and Mellor have been widely received, argue in favor of *facts* to be the relata of singular causation.² Between those two poles there is much room for intermediary positions. Authors such as Bennett, Mackie, Kistler, or Dowe claim that both events and facts can, depending on the context, function as causes and effects.³ Moreover, Vendler holds that effects are events

¹ Cf. Davidson (1980 (1967)), Hausman (1998), pp. 18-28, Spirtes, Glymour, and Scheines (2000 (1993)), p. 20, Ehring (1997), pp. 71-90, Kim (1973), Kim (1993), Suppes (1970), pp. 68 et seq.

² Cf. Horgan (1978), Mellor (1991 (1987)), Mellor (2001).

³ Cf. Bennett (1988), Mackie (1974), ch. 10, Kistler (1999), Dowe (2000), pp. 168-170.

while causes are facts.⁴ Other notions, whose relationship to the polarity of event and fact theories of cause and effect is not all that obvious, have been brought into play. Ehring takes the relata of singular causation to be tropes (or property instances), Salmon presumes causes and effects to be processes or interactions, Aronson and Fair advocate objects as relata of causation, while von Wright favors states of affairs.⁵

Of course, this list of different answers to the question as to what entities are causally related on token level does not intend to provide more than a plain labelling of these positions. Their mutual conceptual relationships are intricate and ask for clarification. In the context of the inquiry at hand there is no room to meet that task. Still, the ontology of cause and effect implemented for the present study shall be developed by comparison of the two most prominent accounts of what the relata of singular causation are: fact and event causation.

Arguments in favor or against fact and event theories of singular causation often proceed from linguistic analysis of causal statements or sentences. This is a rather surprising starting point for such arguments, because nothing conclusive regarding the nature of causes and effects can be learned from natural language. According to Kistler (1999), p. 26, “[i]n English, there exist at least the following types of expressions capable of identifying as a cause of my surprise something about Mary’s performing the song”:

- (1) her performing the song
- (2) the performing of the song
- (3) the performance of the song
- (4) that she performed the song
- (5) the fact that she performed the song.

Vendler (1962) and Bennett (1988) claim that (1), (4) and (5) have factual meaning, whereas (2) and (3) refer to events. The same variety of expressions could be implemented to designate the effect: “my surprise”, “the fact that I was surprised” etc. Both of the following are well-formed English causal statements:

- (6) The performance of the song caused my surprise.
- (7) The fact that she performed the song caused the fact that I was surprised.

This shows either (a) that linguistic analysis cannot answer the question as to what entities are related as cause and effect, because natural language apparently does not favor events or facts in causal contexts, or (b) that causes and effects can be both events and facts.⁶ Authors that do not opt for (b) usually try to show that one of (6) and (7) is primitive and the other dependent on the other. Davidson (1980), for instance, claims that (7) is to be analyzed in terms of (6), i.e. that facts can

⁴ Cf. Vendler (1967).

⁵ Cf. Ehring (1997), ch. 3, Salmon (1977), Aronson (1971), Fair (1979), von Wright (1973),

⁶ Cf. e.g. Bennett (1988) or Mackie (1974).

only be causally related given that corresponding events are so related, whereas e.g. Horgan (1978) argues that facts are the fundamental type of entities capable of interacting causally. This is the dispute we shall be concerned with in the following.

All the arguments for and against fact and event theories are heavily disputed and call for a theoretical framework which in the debate is often not as sharply developed as would be required. The opposition between fact and event theories of causation is not at all as clear-cut as it may prima facie seem. On the contrary, both conceptions depend on a vast theoretical background and they vary according to what kind of entities facts and events are taken to be. Moreover, a widespread blending of the notions of event and fact can be found in the literature. Reichenbach, for instance, proposes to use the notions of fact and event synonymously.⁷ Mellor analyzes facts as property exemplifications by objects at times, which is identical to Kim's analysis of events.⁸ Comparably, Taylor takes events to be a species of facts, which essentially he spells out on par with Mellor. Events, according to Taylor constitute the subclass of facts that involve change.⁹ Or Baylis (1948) contends that facts are particulars, which coincides with Davidson's (1967) view concerning events. Finally, Austin construes facts so broadly that "phenomena, events, situations, states of affairs"¹⁰ all come to be facts.

However, the ongoing debate over fact vs. event causation is not normally perceived to be merely terminological. Controversy can only arise within a theoretical make-up according to which facts and events are not just different labels for one and the same kind of thing, but ontologically differing entities. Before we can thus set out to compare the merits and shortcomings of event and fact causation, we need to settle for some rudimentary understanding of what events and facts shall be taken to be. For the purposes of the context at hand, therefore, the following minimal contrast between events and facts, for which Ramsey (1927)¹¹ has most notably argued, will be presumed: Events are particulars¹² – simple or complex – to which reference is made by means of names and definite descriptions, whereas facts are expressed, represented or stated by sentences, they cannot be referred to by singular terms.¹³ Furthermore, events are dated and have a spatiotemporal locality. Facts, on the other hand, are undated entities that are not located in space

⁷ Cf. Reichenbach (1947), §48. Reichenbach's doctrine that equates facts with events runs into difficulties when confronted with e.g. facts as ' $2 + 2 = 4$ ' (cf. Williamson (1976), p. 207).

⁸ Cf. Mellor (1991 (1987)), pp. 203-204, and Kim (1973).

⁹ Cf. Taylor (1985), ch. 4.

¹⁰ Austin (1970), p. 156.

¹¹ Similarly: Mellor (1991 (1987)), p. 202.

¹² As mentioned above, there are some fact theorists that hold facts to be particulars as well. Representatives of the event camp, on the other hand, widely agree on the categorization of events as particulars (cf. Kanzian (2001), ch. II.3).

¹³ It must be noted that the notion of a definite description is here understood in Russell's terms. That means, a definite description is taken to be of the form ιxFx , or colloquially "the ... such that ... is F ". Thus, an expression as "the fact that ..." does not have the form of a definite description. The 'that'-clause states the corresponding fact, it does not predicate anything of the fact – it does not feature a variable running over facts. Without making his reasons explicit, Mellor (1995), pp. 115-116, rejects this understanding of definite descriptions. He views expressions as "the fact that

and time. The fact that Neil Armstrong stepped onto the surface of the moon on July 20, 1969, is not located on the lunar surface in 1969; there has not been a time before it was a fact, there will not be a time when it is not a fact any more, and there never was and never will be a place where it is not a fact – it simply *is* a fact. In contrast, the first human step onto the surface of the moon is an event which has a spatiotemporal locality: It took place on the lunar surface on July 20, 1969.

The following paragraphs shall outline and evaluate the main arguments in the debate over fact vs. event causation. By highlighting where advantages and disadvantages of fact and event causation lie, the following discussion intends to clarify what requirements a successful analysis of the causal relata has to meet. Based on this exposition the subsequent section will then develop such an analysis for the theoretical needs of the investigation at hand.

Arguments from the Identity of the Causal Relata

Besides further spelling out the differences between facts and events, analyses of these two notions have to meet the requirements of a theory of causation. And here severe problems linger. The three most widely received analyses of the concept of event are deficient in this respect: Davidson's (1980) proposal, which considers events to be unanalyzable particulars individuatable only by their causes and effects, has, especially when implemented into a theory of causation, "an air of circularity"¹⁴. For in order to determine whether *a* and *b* refer to the same event, one has to ascertain the identity of *a*'s and *b*'s causes and effects, which again, according to Davidson, are events whose identity is only establishable by resolving the identity of their causes and effects, and so on ad infinitum.¹⁵

As an alternative, Quine (1985) proposes to treat events on par with physical objects. To this end, he resorts to the so-called Lemmon-criterion, according to which events are identical iff they occupy the same spacetime zone.¹⁶ Of course, spatiotemporal boundaries of events are vague. It is unclear when exactly Sebastian's walk begins and ends or what region in space it occupies. This however, Quine holds, is irrelevant to most of our macroscopic contact with events, i.e. it does not normally hinder the differentiation between macroscopic events. Though the vagueness of the spatiotemporal extension of events might in uncommon cases retard or fully foreclose the application of Quine's criterion, in return, it is free of every "air of circularity", for coordinates in space and time, following Quine, are identifiable independently of events.

However, granted its non-circularity, Quine's criterion suffers from other drawbacks. For instance, there are events that occupy the exact same spacetime zone,

...” to be singular terms referring to facts. (Notwithstanding this referability to facts, however, he does not hold facts to be particulars.)

¹⁴ Davidson (1980 (1969)), p. 179.

¹⁵ Cf. Bennett (1988), ch. 6, see also Quine (1985), p. 166.

¹⁶ Cf. Lemmon (1967).

but that we would not want to identify. A famous example of this sort is Davidson's metal ball that heats up and rotates at the same time.¹⁷ According to Quine, the heating and rotating of this metal ball have to be identified. Prima facie, this consequence is counterintuitive, but for the sake of a straightforward criterion for the identity of events Quine "is not put off by the oddity of such identifications"¹⁸ and adjusts his intuitions to his analysis. Notwithstanding a possible adjustment of intuitions in this vein, an implementation of Quine's criterion into a theory of causation is problematic. Consider a frisbee that rotates just as long as it flies through the air.¹⁹ We would distinguish two events – the rotation, event *a*, and the flight, event *b* – and hold that *a* causes *b*. Nonetheless, against the background of Quine's identity criterion, *a* and *b* have to be identified. Hence, a theory of causation implementing Quine's event notion would have to credit *b* with causing itself. Yet, there are good reasons for denying self-causation the status of an acceptable notion – the relation of singular causation is irreflexiv. Causes are entities that in some sense or other – be it in terms of regularities, probabilistic or counterfactual dependencies – determine their effects. Every effect trivially determines itself, for it is trivially true that whenever the effect *x* is given, *x* is given; or that *x* raises the probability of *x*; or that had *x* not happened, *x* would not have happened. Within the framework of Quine's overly *coarse-grained* event notion, however, self-causation would have to be accepted in order to account for examples of the frisbee type. Such a theoretical framework would collapse all presently known theories of causality into triviality and it would not be able to account for the irreflexivity of singular causation.

Finally, Kim's (1976) account, that presumes events to be property exemplifications by objects at times, has events to be identical iff their constitutive properties, objects, and times coincide. As a result of this criterion, the world is populated by an ample amount of different events that we would intuitively not discriminate. For whenever different constitutive properties are predicated of objects, according to Kim, different events are referred to. This is especially irritating for a theory of causation, because causes and effects – taken to be events – certainly are not proliferated by their mere redescription. Consider the following examples:

- (8) The stabbing of Caesar by Brutus caused Caesar's death.
- (9) The infliction of a wound in Caesar's heart by Brutus caused Caesar's death.

Details concerning identity criteria of predicated properties left aside, it is clear that stabbing someone and inflicting a wound in someone's heart differ – extensionally as well as intensionally. Hence, both of these statements predicate different properties of Brutus' attack on Caesar, and thus, according to Kim's analysis, they identify different causes of Caesar's death. However, we would not say that

¹⁷ Cf. Davidson (1980), pp. 178-179.

¹⁸ Quine (1985), p. 167. Similarly, Xu (1997) gives preference to the straightforwardness of the Lemmon-criterion over its counterintuitive consequences when applied to events (cf. Xu (1997), p. 147-157).

¹⁹ Cf. Meixner (1997), p. 105.

Caesar's death was on the one hand caused by Brutus' stabbing and on the other by Brutus' infliction of a wound. Both of these descriptions refer to the same cause of Caesar's death: Brutus' attack on Caesar in the Forum Romanum on March 15, 44 B.C. Hence, Kim's event analysis is too *fine-grained* to be integratable into a satisfactory theory of causation.

There are many other accounts of event identity available. Some draw on modal notions as Brand's (1977), others, as Lombard's (1979) or Cleland's (1991), analyze event identity with recourse to other unclear – and, above all, causally entrenched – ontological categories as e.g. changes or processes. While none of these accounts has been or can be used as an adequate fundament of a Humean analysis of causation, we shall see in section 2.2.2 that – as e.g. Bennett (1996) argues – the Kimian analysis has good prospects of being thus modified that a proliferation of causes and effects by redescription is prevented.

It is no more straightforward to explicate the ontology of causation with recourse to popular theories of facts. Fine (1982) broadly distinguishes between three types of conceptions of fact: One holds facts to be the truth of a proposition or statement, another identifies facts with true propositions or statements, and the third views facts to “be structured entities or complexes, built up in certain characteristic ways from their constituents”.²⁰ Fine calls the first two conceptions *propositional* and the third *worldly*. Causal intuitions are clear in one respect: Causal dependencies do not subsist between any kind of propositional or linguistic entities. Causes and effects are entities somehow subsisting in nature. They are independent of our linguistic reference to them. Therefore, the first of Fine's categories of fact conceptions is unsuited for a theory of causation.

According to a worldly conception, facts are not identical with, but recorded, expressed or stated by true statements or sentences. Facts are what is the case if statements are true. Several different analyses of the structure of facts have been put forward in the literature. The standard view among these different accounts is what Russell, for instance, expresses thus:

What I call a fact is the sort of thing that is expressed by a whole sentence, not by a single name like ‘Socrates’. (...) We express a fact, for example, when we say that a certain thing has a certain property, or that it has a certain relation to another thing; but the thing which has the property or the relation is not what I call a ‘fact’.²¹

A fact is a thing having a property or, more generally, the subsistence of a relation between particulars. Facts, thus, are somehow constituted of properties or relations on the one hand, and things on the other. This conception can easily be generalized such that the second constituent of facts does not have to be a particular thing, but a

²⁰ Fine (1982), p. 52. Representatives of the first conception are e.g. Ducasse (1940), p. 710, and Carnap (1947). The second conception is e.g. held by Moore (1953). The third can be found in Russell (1977), p. 182, or it has been traced back to Wittgenstein (cf. Lampert (2000), pp. 280 et seq.).

²¹ Russell (1977), pp. 182-183.

point or a region in spacetime. How exactly relations and things or spatiotemporal points have to combine in order to constitute a fact is a question to which many conflicting answers have been proposed in the literature²² – a question, however, that is of no relevance to the following discussion.

It is important to stress again that the relationship between sentences and corresponding facts is not one of reference, rather “the structure of a true sentence mirrors the structure of the fact for which it stands”.²³ Within this theoretical framework, facts are identified by resorting to their structure, that is, facts are identical just in case their structures coincide. When exactly facts structurally coincide is an intricate question, on which more will be said as we proceed. What is accepted by all authors that have addressed this question with respect to worldly facts and that have analyzed facts along the standard lines exemplified by Russell’s quote above, is the following negative condition: Whenever intensionally as well as extensionally differing predicates are implemented to describe objects – or, more generally, spatiotemporal localities – and their mutual relation, a different fact is being expressed or stated.

At this point the theoretical vicinity of such a worldly notion of fact to Kim’s event theory becomes apparent. Consequently, similar objections to building a theory of causation upon a worldly fact conception can be raised as have been put forward against Kim’s analysis of events.

- (10) The fact that Brutus stabbed Caesar caused the fact that Caesar died.
- (11) The fact that Brutus inflicted a wound in Caesar’s heart caused the fact that Caesar died.

Relative to a straightforward logical analysis of (10) and (11) in terms of their grammatical structure, (10) predicates of Brutus “. . . stabbed Caesar”, whereas (11) predicates of Brutus “. . . inflicted a wound in Caesar’s heart”. Against the background of the negative condition for fact identity mentioned above, (10) and (11) single out different facts as causes of Caesar’s death. Yet, intuitively we would say that, even though (10) and (11) employ different predicates, they represent the same cause – under different descriptions – of Caesar’s death. It is Brutus’ action on March 15, 44 B.C. that caused the end of Caesar’s life, regardless of the way this action is described. By describing causes in various ways they are not multiplied. Fact theories, just as Kim’s event theory, thus yield an overly fine-grained analysis of the causal relata.

The fact theorist could reply that a logical analysis of (10) and (11), that hinges on the grammatical structure of these sentences, is oversimplifying matters. Whether fact reporting expressions state identical facts can, according to e.g.

²² For one possible answer see Lampert (2000), ch. 5.

²³ Neale (2001), p. 204.

Wittgenstein, only be determined by resorting to a standardized language of elementary sentences. Elementary sentences in Wittgenstein's terminology are sentences predicating specific colourings of spacetime points.²⁴ Wittgenstein holds that facts are individuatable only on such a level of specification. The representative of fact causation could therefore claim that a proper analysis of (10) and (11) in terms of Wittgensteinian elementary sentences would reveal that these two sentences, contrary to grammatical appearances, state identical facts after all. This, however, would be an entirely hypothetical claim, for presently there are no criteria available with respect to how a reduction of natural language to a standardized language of elementary sentences could be accomplished. In fact, it could be argued that the question as to what expression within the language of elementary sentences corresponds to (10) and (11) presupposes an answer to the question whether (10) and (11) state identical facts or not. Whatever the presuppositions of a successful analysis of (10) and (11) in terms of elementary sentences might be, it is clear that a friend of fact causation trying to account for our coarse-grained causal intuitions in this Wittgensteinian vein would have to provide standards of reducing natural language to elementary sentences – indubitably an extremely delicate task –, before he could even begin to think about causation.

All in all, these arguments from the identity of the causal relata compromise both event and fact causation and, thus, do not prefer one over the other. They demonstrate that there is a gap between two theoretical concerns: on the one hand the analysis of events and facts, and on the other hand the analysis of the causal relata in a way that can serve as a basis for an adequate theory of causation. Without further adjustments existing event and fact analyses are unsuited for an efficacious theoretical reproduction of common causal judgements as, for instance, concerning the number of causes of Caesar's death. Several adjustments are conceivable. For instance, events, considered to be unanalyzable particulars in Davidson's sense, whose identity is independent of linguistic reference to them, would perfectly well account for coarse-grained causal judgments, provided their identity could be accounted for in a non-circular way. Moreover, a conflict between common causal judgments and a theory of causation based on e.g. Kimian events or worldly facts is not necessarily decisive against the theory, it could also be taken as grounds on which to revise common causal judgements.²⁵

The Slingshot-Argument

An argument often raised – most prominently by Davidson²⁶ – against fact theories of causation aims to show first that the sentential connective "... causes/d ..." linking fact-reporting expressions is a truth-function, which in fact it cannot be, and

²⁴ Cf. Lampert (2000), ch. 5.

²⁵ Taylor, for instance, is ready to abstain from contrary intuitions and accept an analysis of events and facts that is overstrict in the way it counts its analysans (cf. Taylor (1985), p. 90).

²⁶ Cf. Davidson (1980), pp. 151-153. In a similar vein, this argument has been put forward by Church (1943) or Quine (1980b). See also Williamson (1976), p. 210, Mackie (1974), ch. 10.

second that, if facts are taken to be token level causes, any effect is caused by any fact. Barwise and Perry (1996) have labeled this argument the *slingshot-argument*. The literature treating the slingshot-argument and its consequences is vast. By far the most thorough reconstruction of the slingshot can be found in Neale (2001). Neale's valuable recent work has evoked an ongoing discussion on the topic. The slingshot-argument does not only affect the analysis of the causal relata as facts, but philosophical recourse to facts in general. These broad philosophical implications, however, cannot be given adequate consideration in the context at hand. I will therefore confine myself to versions of the slingshot that, from the outset, are tailored to the requirements of the analysis of the relata of singular causation.

The slingshot-argument, as presented by Davidson, proceeds from two inference principles: The truth-value of causal statements as (10) or (11) is unaltered (i) if expressions reporting the causing fact or the caused fact are substituted by logically equivalent expressions, and (ii) if singular terms are replaced by others referring to the same particulars. Both (i) and (ii) are standard inference principles of extensional logic. (i) expresses the *Principle of Substitutivity for Logical Equivalents*, PSLE for short. Technically put, PSLE says that if for two statements ϕ and ψ : $\phi \models \equiv \psi$, then, if $\Sigma(\phi)$ is a true sentence containing at least one occurrence of ϕ , the sentence $\Sigma(\psi)$ is also true, where $\Sigma(\psi)$ results from replacing at least one occurrence of ϕ in $\Sigma(\phi)$ by ψ . In contrast, (ii) states the *Principle of Substitutivity of Singular Terms* (PSST). PSST maintains that if two singular terms α and β have the same referent, i.e. $\alpha = \beta$, then, if the sentence $\Sigma(\alpha)$ containing at least one occurrence of α is true, $\Sigma(\beta)$, which results from $\Sigma(\alpha)$ by replacing at least one occurrence of α in $\Sigma(\alpha)$ by β , is true as well. The validity of both PSLE and PSST is a necessary condition for a sentential context to be extensional, i.e. truth-functional. A sentential context is truth-functional iff material equivalents are substitutable *salva veritate* (s.v.), the latter being the *Principle of Substitutivity of Material Equivalents* (PSME). The validity of PSME for a sentential context implies the validity of both PSLE and PSST. Therefore, if for a sentential context either PSLE or PSST is not valid, PSME does not hold for this context, which therefore must be intensional.

Given the intuition that causes and effects are entities in nature that are independent of our linguistic reference to them, the validity of PSLE and PSST seems plausible, to say the least. Logically equivalent statements, i.e. statements that agree in truth-value under all systematic reinterpretations of their non-logical vocabulary, do not seem to be reporting different facts. For whatever is the case if the first of these statements is true, is necessarily likewise the case if the second statement is true.²⁷ Analogously for singular terms: What is the case given that a statement reporting the exemplification of a property by a particular a is clearly

²⁷ Correspondingly, Prior (1948), p. 62, for instance, holds that the logical equivalence of two statements is a sufficient condition to the effect that, if one of the statements expresses a fact – only true statements express facts –, both statements express the *same* fact: “‘All Mammals are vertebrates’” and ‘All non-vertebrates are non-mammals’ are different but equivalent propositions, and express (...) the same fact”.

unaltered by exchanging singular terms used to designate a . Therefore, *prima facie* there are no objections to the acceptance of PSLE and PSST for statements of type (10) and (11). Moreover, consider the following statements:

- (12) The fact that the eldest son of George Bush senior fell off the ladder caused the fact that George W. Bush broke his leg.
- (13) The fact that the 43rd president of the USA fell off the ladder caused the fact that George W. Bush broke his leg.
- (14) The fact that the 43rd president of the USA did not fall off the ladder caused the fact that George W. Bush broke his leg.

If a theory of fact causation draws on mirroring our coarse-grained causal intuitions to the effect that (12), (13), and (14) express the same cause of the fact that George W. Bush broke his leg, there are good theoretical reasons to accept the validity of PSLE and PSST for factual causal statements.

Given the validity of PSLE and PSST and the standard definition of the universal class $\{x : x = x\}$ along with the notion of logical equivalence borrowed from the standard set calculus, the Davidsonian slingshot-argument runs as follows:

- | | | | |
|-----|-----|---|---------|
| 1 | [1] | $p \equiv q$ | A |
| 2 | [2] | The fact that p caused the fact that z . | A |
| 2 | [3] | The fact that $\{x : x = x \wedge p\} = \{x : x = x\}$ caused the fact that z . | 2, PSLE |
| 1,2 | [4] | The fact that $\{x : x = x \wedge q\} = \{x : x = x\}$ caused the fact that z . | 3, PSST |
| 1,2 | [5] | The fact that q caused the fact that z . | 4, PSLE |

$\{x : x = x \wedge p\}$ refers to the universal class iff p is true and to the empty class iff p is false. Hence, $\{x : x = x \wedge p\} = \{x : x = x\}$ is true whenever p is true and false whenever p is false. $\{x : x = x \wedge q\}$, in turn, refers to the universal class iff q is true and to the empty class iff q is false. Since p and q are materially equivalent (cf. line [1]), $\{x : x = x \wedge q\}$ designates the universal class just in case $\{x : x = x \wedge p\}$ does so too and the empty class whenever $\{x : x = x \wedge p\}$ refers to the empty class as well. Thus, $\{x : x = x \wedge p\}$ and $\{x : x = x \wedge q\}$ are co-referring singular terms.

The slingshot-argument shows two things: First, if PSLE and PSST are accepted, not only logically, but also *materially* equivalent expressions are substitutable *s.v.* in factual causal statements, i.e. PSME also holds for factual causal statements. This means that these statements are truth-functional. Second, any fact is just as much the cause of the fact that z as any other, no matter what the fact that z is. Both of these conclusions are unacceptable. Replacing a true statement expressing a causing fact by *any* other true statement does not necessarily retain the truth-value of the complex causal statement.

These unwanted consequences have evoked all kinds of different reactions. The event camp, as already mentioned, has accepted the consequences of the slingshot and has used the argument to show that facts cannot be causes and effects. The slingshot-argument cannot be raised against event causation, at least not against event causation based on a notion of event – as introduced above (cf. p. 23) – that takes events to be particulars to which reference is made by singular terms. For, generally, the slingshot is aimed at theories that hold *statements* or *sentences* to state or by some means stand for extra-linguistic entities. However, event causation, as it is ordinarily understood, does not treat “...causes/d ...” as a sentential connective, but as an ordinary two-place relation. The question as to the truth-functionality of “...causes/d ...” therefore does not emerge in the first place. Furthermore, whenever within ordinary relations co-referring singular terms are substituted, the overall truth-value certainly does not change.

The fact camp, in contrast, cannot welcome the consequences of the slingshot and has undertaken considerable efforts to rebut the argument. The objections raised against it can be broadly categorized in two groups: On the one hand, there are those – the majority of fact theorists – who deny one or both of the inference principles PSLE and PSST for factual causal contexts or factual contexts in general, and, on the other hand, there are some who deny the logical equivalence of p and $\{x : x = x \wedge p\} = \{x : x = x\}$ – the co-referentiality of $\{x : x = x \wedge p\}$ and $\{x : x = x \wedge q\}$, given $p \equiv q$, being undisputed.

As a representative of the first category, Mellor (1987), for instance, refuses the applicability of PSST to factual causal statements.²⁸ He claims that it is a necessary condition for one fact to be the cause of another that the first fact raises the probability of the second, and, since “probabilities are not truth functions”²⁹, substituting co-referring singular terms may change the probabilities of corresponding facts. According to Mellor, e.g. the fact that the 43rd president of the USA enters the Oval Office does not have the same probability as the fact that the eldest son of George Bush senior enters the Oval Office. This claim, of course, only goes through when some intensional notion of probability is implemented. As regards an extensional relative frequency account of probability, exchanging co-referring singular terms could not change probability values. The relative frequency of entrances to the Oval Office by the 43rd president of the USA is identical to the relative frequency of entrances to the Oval Office by the eldest son of George Bush senior.

For Barwise and Perry (1996) it is PSLE which is unacceptable.³⁰ They analyze sentences as singular terms that refer to situations. Against this background, logically equivalent expressions are only substitutable *s.v.* if they refer to the same situation. And, following Barwise and Perry, it takes much more for the co-referentiality of two sentences than logical equivalence – co-referring sentences

²⁸ This strategy to block the slingshot can also be found in Bennett (1988), §16.

²⁹ Mellor (1991 (1987)), p. 212.

³⁰ Barwise and Perry (1996), p. 380. In a similar vein Anscombe (1969) and Koons (2000) reject the validity of PSLE for causal sentences.

implement extensionally as well as intensionally identical predicates and their context of utterance coincides.

The objections to the slingshot put forward by Mellor, Barwise, and Perry result in the opacity of factual causal statements. This not only blocks the slingshot-argument, but, in addition, directly leads to the desired non-truth-functionality of causal statements that relate causing and caused facts. Yet, denying the simultaneous validity of PSLE and PSST is a high price to pay. PSLE and PSST are very much backed by ordinary causal judgements. Causal dependencies commonly are not taken to subsist subject to the expressions by which they are stated or referred to. Representatives of fact causation usually do not want to deny this, or more specifically, they do not want to deny that sentences expressing causes or effects are substitutable *s.v.* within causal sentences *if they state identical facts*. However, so they continue, it is not the case that two fact reporting expressions p and q state the same fact iff p results from q , and vice versa, by a suitable application of PSLE and PSST. Mutual convertibility through PSLE and PSST, the slingshot-argument shows, cannot be sufficient for fact identity. Instead, e.g. Williamson (1976) proposes to resort to Leibnizian identity criteria: Two facts are identical iff they share all properties. How this conception is supposed to block the step from line [2] to line [3] Williamson illustrates as follows:

(...) it may be argued that the fact that $[p]$, where $['p']$ is an abbreviation for some proposition, is never identical with the fact that $[\{x : x = x \wedge p\} = \{x : x = x\}]$ ³¹, since, among other things, the second fact concerns sets while the first does not; the second is the fact that one thing is identical with another, while the first is not; the second is arcane, while the first is widely known, and so on.³²

Spelling out fact identity in this vein – besides interpreting causal statements to be opaque – yields an even more fine-grained analysis of the causal relata than a conception of fact identity along the lines of PSLE and PSST. Whenever “the fact that p ” and “the fact that q ” do not express corresponding facts by intensionally coinciding predicates, they represent different causes or effects. Analyzing fact identity along these lines in the end comes down to individuating facts by means of true statements. This, in turn, would deprive facts of their intended function as truth-makers of statements, since in view of their individuation in terms of true statements they could no longer be referred to in order to resolve the truth of statements.³³ Needless to say that counterintuitivity and problems arising for facts as truth-makers of statements are no coercive objections against these ways around the unwanted consequences of the slingshot-argument. A theory of fact causation does not have to draw on analyzing truth in terms of facts, nor does it necessarily have to keep in line with causal intuitions. Still, these considerations show

³¹ Williamson uses the set abstraction operator \hat{x} to denote sets. Therefore, his notation – as well as Cummins’ and Gottlieb’s notation below – had to be accommodated to the one at hand.

³² Williamson (1976), p. 211.

³³ Cf. Strawson (1999), pp. 167-168.

that Mellor's, Barwise's, and Perry's strategies to immunize fact causation against the slingshot have extensive consequences in other philosophical areas or require additional theoretical work to be done within these other areas.

Refutations of the slingshot that do not not introduce intensionality and/or an increase of the number of causes for each effect and nevertheless lead the way for fact causation all fall into the second category mentioned above. Cummins and Gottlieb (1972), for instance, argue against the logical equivalence of p and $\{x : x = x \wedge p\} = \{x : x = x\}$:

[I]n first order logic with identity, "... $c \dots \supset (\exists x)(x = c)$ " is provable, where ' c ' is any expression which can occupy the position of an individual variable. Thus, $[\{x : x = x \wedge p\} = \{x : x = x\}]$ logically implies the existence of [the universal set] no matter what sentence p represents. But then that sentence is not logically equivalent to the sentence $[\{x : x = x \wedge p\} = \{x : x = x\}]$, since it is not the case that every sentence implies the existence of [the universal set].³⁴

Mackie (1974) has rightly pointed out that this objection to the slingshot-argument is not compelling, for the definition of the universal class is an axiom of the set calculus implemented for the Davidsonian slingshot. Therefore, contrary to Cummins and Gottlieb, $\{x : x = x \wedge p\} = \{x : x = x\}$ is straightforwardly derivable from p within the set calculus.³⁵

The question as to whether p and $\{x : x = x \wedge p\} = \{x : x = x\}$ are logically equivalent in the end comes down to the question as to what notion of logical equivalence is to be employed for factual causal statements. Logical equivalence always has to be understood relative to some formalism. Given the notion of logical equivalence of the standard set calculus, it is hardly questionable that p and $\{x : x = x \wedge p\} = \{x : x = x\}$ are logically equivalent. What might be questionable, though, is whether it is adequate to borrow the notion of logical equivalence from the set calculus for factual causal statements. p represents any statement, say

(S₁) Brutus stabbed Caesar.

Then, $\{x : x = x \wedge p\} = \{x : x = x\}$ stands for the statement

(S₂) The class such that its members are identical to themselves and Brutus stabbed Caesar is identical to the class such that its members are identical to themselves.

If the truth of (S₂) is taken only to depend on the truth of (S₁) and vice versa, (S₁) and (S₂) are informally judged to have the same truth-values under all systematic reinterpretations of their non-logical vocabulary. (S₁) and (S₂) are thus seen to be

³⁴ Cummins and Gottlieb (1972), p. 265.

³⁵ Cf. Mackie (1974), p. 252.

informally equivalent.³⁶ That means, the set theoretic supplement in (S₂) is redundant – irrelevant to the truth-conditions of (S₂). (S₂) does neither state more nor less than (S₁). It could therefore be contended that both statements are to be formalized either by p or by $\{x : x = x \wedge p\} = \{x : x = x\}$. Yet, a formalization of (S₂) by p , which would enable a simple recourse to logical equivalence of propositional logic, would block the step from line [3] to line [4]. It might thus be argued that a set theoretic formalization of (S₂) by $\{x : x = x \wedge p\} = \{x : x = x\}$ is *prima facie* not motivated, or at least would require some additional motivation by the proponents of the slingshot-argument.

The latter will retort that there are conventions of logical formalization requiring an adequate formalization to reveal as much of the internal structure of the formalized statement as possible, for only thus all of the logical dependencies between colloquial statements become reproducible by logical formalisms.³⁷ Therefore, a formalization of (S₂) by $\{x : x = x \wedge p\} = \{x : x = x\}$ ought to be considered adequate. This turns the burden of vindication back over to the opponent of the slingshot. The fact theorist now has to provide formalization standards that generally ask for minimal formalizations, i.e. for formalizations that only capture truth-relevant features of formalized statements. According to such standards, all informally equivalent statements would have to be formalized by the *same* (minimal) formula, which, in turn, would deprive logic of the task of proving equivalencies.

However, the opponent of the slingshot does not necessarily have to go as far as to generally ask for minimal formalizations. Instead of requiring a confinement to propositional logic, he might move towards the proponent of the slingshot in conceding that more of the internal structure of (S₂) ought to be revealed than is possible by means of propositional logic alone. For instance, he might maintain that class identity should be defined contextually, i.e. in terms of (2.1):

$$\{x : Fx\} = \{x : Gx\} =_{df} \forall x(Fx \equiv Gx) \quad (2.1)$$

$$\{x : x = x \wedge p\} = \{x : x = x\} \quad (2.2)$$

$$\forall x(x = x \wedge p \equiv x = x) \quad (2.3)$$

Applied to the slingshot, definition (2.1) yields that (2.2) is to be understood in terms of (2.3). If instead of (2.2) one substitutes (2.3) for p on line [3], the slingshot will not proceed beyond that line, for PSST cannot be applied to (2.3). Hence, the opponent of the slingshot proposes to borrow the notion of logical equivalence from standard first-order logic with identity, and not, as does Davidson, from the

³⁶ *Informal equivalence* which can be predicated of natural language statements must be clearly distinguished from *formal equivalence* which is a notion that applies to formulas and stands for equivalence relative to some calculus. Thus, formal equivalence is what in the present context has so far been and will subsequently be called *logical equivalence*. For details on these notions cf. e.g. Brun (2003), pp. 37-40.

³⁷ For details concerning criteria of adequate formalization cf. Baumgartner and Lampert (forthcoming).

set calculus.³⁸ What about the merits of this argument? First of all, the proponent of the slingshot will insist that (2.2) is closer to the internal structure of (S₂) than (2.3) and, therefore, constitutes a more adequate formalization of (S₂) than (2.3). Secondly, (2.1) guarantees for the substitutability of (2.2) and (2.3). The proponent of the slingshot, once the latter's opponent has substituted (2.3) for (2.2), hence will just reverse this substitution on grounds of (2.1) and run the slingshot anew. He can justify his insistence on the implementation of the set calculus with standard arguments regarding the preferability of richer formalisms for reasons of their ability to reveal more logical dependencies. Eventually, the opponent of the slingshot will be forced to either develop formalization standards that explicitly prohibit a formalization of (S₂) by (2.2), or he has to renounce the validity of either PSLE or PSST. He cannot at same time maintain a contextual definition of class identity along the lines of (2.1) and the validity of both PSLE and PSST.

There is another way to refute the slingshot-argument by rendering the formalization of statements such as (S₁) and (S₂) dubious. Contrary to what has been said above, it could be held that (S₁) and (S₂) are *not* informally equivalent after all. Then, every formalization of (S₁) and (S₂) that results in logical equivalence of the corresponding formulas could be rejected on grounds of being inadequate. Again, the slingshot-argument would not proceed beyond line [3]. How could the informal inequivalence of (S₁) and (S₂) be substantiated? By predicating a property of an object (or, more specifically, a person) (S₁) unquestionably states a worldly fact. In contrast, (S₂) is a statement about the alleged identity of two classes and e.g. Wittgenstein, as is well known, plainly denies that identity statements of type (S₂) are meaningful and professes to eliminate “=” from logical formalisms, which, instead, shall be exclusively interpreted.³⁹ At best, according to Wittgenstein, identity can meaningfully be predicated of co-referring singular terms, in which case identity statements are to be read as rules that allow for the mutual substitution of the two terms connected by “=”.⁴⁰ Thus, with recourse to Wittgenstein the fact theorist could back up his rejection of the informal equivalence of (S₁) and (S₂) somehow along the following lines: (S₁) states a worldly fact, whereas (S₂) is either meaningless or is a rule that allows for the substitution of the expressions on both sides of the identity predicate. In any case, it does not state a worldly fact and hence it cannot be informally equivalent to (S₁).

Clearly, if Wittgenstein's elimination of the identity sign from logical formalisms is accepted, the slingshot is immediately rebutted for its use of “=” – independently of whether (S₁) and (S₂) are taken to be informally equivalent or not. However, the friend of fact causation taking this route around the slingshot, would have to substantially complete Wittgenstein's sparse suggestions concerning an exclusively interpreted logical formalism that is of acceptable strength, before he could start to think about a logical analysis of causal statements – and this task

³⁸ This essentially is one of the strategies Cummins and Gottlieb (1972) propose in order to refute the slingshot.

³⁹ Cf. Wittgenstein (1995), §§5.5303-5.531.

⁴⁰ Cf. Wittgenstein (1995), §§4.241-4.243, 6.23.

could easily turn out to be highly entangled.⁴¹ Thus, assuming the fact theorist not wanting to follow Wittgenstein as regards the complete elimination of identity from the logical formalism, what about the meaninglessness of identity statements? Authors such as Russell or Quine hold that there are many meaningful (implicit) identity statements as e.g. “Edward VII is the King”⁴² or “The hiding place is known to Ralph and only him”⁴³. More generally, (binary) relations are normally conceived as consisting of pairs of objects, and the identity relation comprises all and only those pairs of type $\langle x, x \rangle$. It is, in fact, hard to see why such a notion should be meaningless.

Finally, suppose that the fact theorist thereupon does not join Wittgenstein in proclaiming the meaninglessness of identity statements. Can he still argue that (S_1) and (S_2) are not informally equivalent on grounds that (S_2) is a terminological substitution rule or, rather, a statement about the substitutability of two singular terms, which (S_1) is not? Strictly speaking, (S_2) can hardly be considered a statement expressing the substitutability of two terms, for no terms are even mentioned in (S_2) . Yet, notwithstanding this restraint and thus granted that (S_2) is a statement about two singular terms, some additional argument would be required as to why the latter alleged feature of (S_2) should foreclose the informal equivalence of (S_1) and (S_2) . All that is needed for informal equivalence is that the two statements have coinciding truth-conditions, and – provided that (S_2) is not meaningless – this seems to be the case for (S_1) and (S_2) , regardless of the fact that (S_2) might be taken to be a statement about singular terms while (S_1) speaks about Brutus. (S_1) and (S_2) are both true iff Brutus stabbed Caesar.

Hence, the informal equivalence of (S_1) and (S_2) might very well be upheld against the challenge of the fact theorist, such that a formalization of (S_1) and (S_2) that results in their logical equivalence can be considered adequate. The debate in the end boils down to the question of the informal equivalence of (S_1) and (S_2) . If this question is positively answered, the slingshot goes through, if it is negatively answered, the argument can be blocked at line [3]. Unfortunately, questions concerning informal equivalencies cannot conclusively be decided argumentatively. Nevertheless, the slingshot indubitably imposes serious theoretical requirements upon the friend of fact causation that opts for this second strategy to block the slingshot. He either has to come up with a theory concerning the adequacy of formalizations that ultimately assigns one and the same (minimal) formula to all informally equivalent statements, or he has to hold that identity statements are meaningless or even that “=” is to be eliminated from logical formalisms.

The slingshot-argument, as presented so far, can be challenged on account of its recourse to the set theoretic notion of logical equivalence, which allows for a formalization of statements of type (S_1) or (S_2) by a formula that introduces a set theoretic surplus which is irrelevant to the truth-conditions of (S_1) or (S_2) .

⁴¹ Cf. Hintikka (1956).

⁴² Russell (1992 (1937)), §64.

⁴³ Quine (1987), p. 91.

However, there is another version of the argument that does not resort to the set calculus. Neale (2001) attributes this second version to Gödel (1951). Gödel's slingshot only uses first-order definable definite descriptions and replaces PSLE by a much more restricted substitution rule, no longer allowing for a substitution *s.v.* of any logical equivalents, but only of a certain form of logical equivalents – so-called *Gödelian equivalents*. Russell's famous analysis of definite descriptions allows for a rephrasing of sentences like Fa by “ a is the x such that $x = a$ and Fx ” or, more colloquially, by “ a is the thing which is F ”. That means, the Russellian theory of descriptions ensures the logical equivalence of (2.4) and (2.5):

$$Fa \quad (2.4)$$

$$a = \iota x(x = a \wedge Fx) \quad (2.5)$$

The *iota*-operator governed (2.5), according to Russell, is nothing but a shorthand for the first-order formula (2.6).

$$\exists x(\forall y((y = a \wedge Fy) \equiv y = x) \wedge x = a) \quad (2.6)$$

This Russellian quantificational interpretation of formulas containing definite descriptions will be of considerable importance upon the philosophical consequences to be drawn from Gödel's slingshot. But for now, let us proceed with Gödel's argument. (2.4) and (2.6) are logically equivalent relative to standard first-order logic with identity. It is thus (2.6) that guarantees for the logical equivalence of (2.4) and (2.5). Gödel says that (2.4) and (2.5) “mean the same thing”.⁴⁴ Moreover, other than, say, $\{x : x = x \wedge Fa\} = \{x : x = x\}$, which is logically equivalent to (2.4) as well, (2.5) does not express a set theoretic surplus that is irrelevant to its truth-conditions. (2.5), just like (2.4), does not speak about anything else than the particular a , and it says nothing over and above a being F . Hence, (2.4) and (2.5) – even against the background of Leibnizian identity criteria for facts – state the same fact, provided, of course, that one of them states a fact. Even if logical equivalents might not be substitutable *s.v.* in factual contexts, expressions of type (2.4) and (2.5) clearly are thus substitutable. Neale labels this substitution rule ι -CONVERSION, abbreviated by ι -CONV. Furthermore, analogously to PSST of the Davidsonian slingshot, Gödel's argument presupposes the substitutability *s.v.* of definite descriptions and names referring to the same particular. That is, whenever for any two definite descriptions $\iota x\phi$ and $\iota x\psi$: $\iota x\phi = \iota x\psi$, then, $\iota x\phi$ and $\iota x\psi$ are substitutable *s.v.* in factual contexts. Likewise, whenever for any definite description $\iota x\phi$ and any name α : $\iota x\phi = \alpha$, then, $\iota x\phi$ and α are substitutable *s.v.* In this regard, Neale speaks of ι -SUBSTITUTION, or ι -SUBS for short.

Given ι -CONV and ι -SUBS, the Gödelian type slingshot – as reconstructed by Neale – runs as follows:⁴⁵

⁴⁴ Gödel (1951), p. 129.

⁴⁵ Cf. Neale (2001), pp. 183-184. In its original form the Gödelian slingshot is not tailored to the causal connective. For the present inquiry though nothing more is of interest. That is why we will confine ourselves to the factual context governed by “... causes/d ...”.

1	[1]	Fa	A
2	[2]	$a \neq b$	A
3	[3]	Gb	A
1	[4]	$a = \iota x(x = a \wedge Fx)$	1, ι -CONV
2	[5]	$a = \iota x(x = a \wedge x \neq b)$	2, ι -CONV
2	[6]	$b = \iota x(x = b \wedge x \neq a)$	2, ι -CONV
3	[7]	$b = \iota x(x = b \wedge Gx)$	3, ι -CONV
1,2	[8]	$\iota x(x = a \wedge Fx) = \iota x(x = a \wedge x \neq b)$	4,5, ι -SUBS
2,3	[9]	$\iota x(x = b \wedge Gx) = \iota x(x = b \wedge x \neq a)$	6,7, ι -SUBS
10	[10]	The fact that Fa caused the fact that p .	A
10	[11]	The fact that $a = \iota x(x = a \wedge Fx)$ caused the fact that p .	10, ι -CONV
1,2,10	[12]	The fact that $a = \iota x(x = a \wedge x \neq b)$ caused the fact that p .	11,8, ι -CONV
1,2,10	[13]	The fact that $a \neq b$ caused the fact that p .	12, ι -CONV
1,2,10	[14]	The fact that $b = \iota x(x = b \wedge x \neq a)$ caused the fact that p .	13, ι -CONV
1,2,3,10	[15]	The fact that $b = \iota x(x = b \wedge Gx)$ caused the fact that p .	14,9, ι -SUBS
1,2,3,10	[16]	The fact that Gb caused the fact that p .	15, ι -CONV

The conclusion of the Gödelian slingshot essentially is identical to what follows from Davidson's argument. Given that ι -CONV and ι -SUBS are valid for factual contexts, as the one of line [10], it follows that PSME is valid for such contexts and, thus, that the latter are truth-functional, moreover, that any fact caused any other fact. What resorts are open to the friend of fact causation in view of Gödel's version of the slingshot-argument? In light of the indisputable first-order logical equivalence of (2.4) and (2.6), the logical equivalence of (2.4) and (2.5) is hardly challengeable without rejecting Russell's theory of descriptions, which, after all, has proven very valuable as regards e.g. empty predications or true negative existential statements. Such a rejection, accordingly, would be a high price to pay.

(S_{2.4}) Brutus stabbed Caesar.

(S_{2.5}) Brutus is the thing (or person) which stabbed Caesar.

(S_{2.6}) Everything that is Brutus stabbed Caesar and there is exactly one Brutus.

In the same vein, denying the informal equivalence of sentences as (S_{2.4}), (S_{2.5}), and (S_{2.6}) – analogously to the aforementioned possible way around the Davidsonian slingshot – and thus denying the adequacy of the formalizations of (S_{2.4}), (S_{2.5}), and (S_{2.6}) by (2.4), (2.5), and (2.6), respectively, would, besides requiring yet to be developed minimal formalization standards, in the end lead to a rejection of Russell's theory of descriptions. Representatives of fact causation not willing to

dismiss the theory of descriptions cannot but concede that expressions as (2.4), (2.5), and (2.6) or sentences as (S_{2.4}), (S_{2.5}), and (S_{2.6}) state the same, i.e. whenever they state a fact, they state the same fact.

A fact theorist neither wanting to give up the theory of descriptions nor to identify facts by means of true statements⁴⁶ has to reject either the applicability of γ -CONV or of γ -SUBS to factual causal statements.⁴⁷ Prima facie, it appears that the fact theorist thereby resorts to the intensionality of causal statements. For, as γ -CONV and γ -SUBS seem to be nothing but special cases of PSLE and PSST it follows that, when either γ -CONV or γ -SUBS is invalid for sentences of type (10) and (11), PSLE or PSST is invalid for these sentences as well. Yet, Neale (2001) points out that this conclusion would be rash. More specifically, he shows that by rejecting γ -SUBS the fact theorist does not automatically have to reject PSST, i.e. the substitutability *s.v.* of co-referring singular terms.⁴⁸ According to Russell's theory of descriptions, definite descriptions are 'incomplete symbols' that do not refer to anything, and, hence, questions as to the co-referentiality of definite descriptions do not arise in the first place. In fact, definite descriptions, on Russell's account, *are not* singular terms.⁴⁹ Definite descriptions never occur in isolation, but only in broader sentential contexts, where, according to Russell, they get a quantificational and *not* a referential interpretation. Expressions of the form "the *F*" are quantificational noun phrases like those of the forms "every *F*" or "some *F*". Or illustrating Russell's point with an example: While (2.7) is a singular predication that states of the particular *a* that it is *F*, (2.8) does not speak of a specific particular, but is a general predication to be understood in terms of (2.9).

$$Fa \quad (2.7)$$

$$F\gamma x(x = a) \quad (2.8)$$

$$\exists x(\forall y(y = a \equiv y = x) \wedge Fx) \quad (2.9)$$

Carried over to the factual context at hand, despite their indubitable close connection, (2.7) and (2.8), following Russell, do not state the same fact, for in Russell's terminology (2.7) expresses a *particular fact*, whereas (2.8) states a *general fact*.⁵⁰ This means, if the friend of fact causation, on par with Russell, treats definite descriptions as non-referring incomplete symbols that only appear in sentences that are to be read as quantificational expressions like (2.9), he can reject the validity of γ -SUBS for factual contexts, without thereby rejecting the substitutability *s.v.* of co-referring singular terms. To him γ -SUBS is wrong, because definite descriptions do not refer in the first place, and, therefore, there cannot be any co-referring definite descriptions.

⁴⁶ Cf. p. 32 above.

⁴⁷ E.g. Oppy (2004) claims that no acceptable theory of facts should take both γ -CONV and γ -SUBS to be applicable to factual contexts, especially to the fact-identity connective.

⁴⁸ Neale (2001), p. 167.

⁴⁹ Cf. Russell (1977), pp. 244-246, 253-254.

⁵⁰ Cf. Russell (1977), pp. 183-184, 234-235.

However, conceding that a quantificational understanding of definite descriptions along with the rejection of ι -SUBS allows the representative of fact causation to rebut the Gödelian slingshot and, at the same time, stick to PSLE and PSST, nevertheless, the consequences of rejecting ι -SUBS for factual causal statements – to a large extent – are the same as the consequences of a refutation of PSST: an overly fine-grained analysis of causal processes. For consider anew statements (12) and (13) on page 30 above. Russell would rephrase the causing fact mentioned in (12) as “The unique thing that is the eldest son of George Bush senior fell off the ladder” and formalize it by (2.10) or equivalently by (2.11), where F stands for “... is the eldest son of George Bush senior” and G represents “... fell off the ladder”.

$$G\iota xFx \quad (2.10)$$

$$\exists x(\forall y(Fy \equiv y = x) \wedge Gx) \quad (2.11)$$

In contrast, the causing fact specified in (13) is to be understood as “The unique thing that is the 43rd president of the USA fell off the ladder” and corresponds to (2.12) and (2.13), where H is “... is the 43rd president of the USA” and G again stands for “... fell off the ladder”.

$$G\iota xHx \quad (2.12)$$

$$\exists x(\forall y(Hy \equiv y = x) \wedge Gx) \quad (2.13)$$

(2.11) and (2.13) are not logically equivalent. Both statements predicate different properties of the particulars satisfying them. They express different facts and, moreover, identify different facts as causes of the fact that George W. Bush broke his leg. Hence, even though a Russellian understanding of definite descriptions might at the same time keep the fact theorist in line with PSLE and PSST and protect him from the, in his view, unacceptable consequences of the slingshot-argument, the invalidity of ι -SUBS will nonetheless render fact causation much more fine-grained than common causal judgements. Within the framework of a Russellian quantificational account of definite descriptions, causes could be multiplied by mere redescription.

Besides, the fact theorist’s rejection of ι -SUBS raises fundamental doubts against the predicate calculus with identity. ι -SUBS can be seen as a mere adaptation of *identity elimination* (=E) for ι -governed expressions. In this regard, ι -SUBS is a purely formal rule that allows for mutual substitutions of expressions of the form $\iota x\phi$ and $\iota x\psi$ or α , provided that a corresponding derivation contains the line $\iota x\phi = \iota x\psi$ or the line $\iota x\phi = \alpha$. Thus understood, ι -SUBS does not put forward any constraints concerning the semantic properties of the expressions on both sides of the identity sign. Russell’s theory of descriptions accounts for the rephraseability or reformalizability of any proper name in terms of a definite description. Now, if =E is not applicable whenever definite descriptions appear on either side of the identity sign, there seems to be no grounds left on which to apply =E to identity statements that purely involve proper names. This, in turn, leads either to a basic

skepticism against the validity of $=E$ in particular and against the meaningfulness of the identity predicate in general, or to a complete elimination of proper names and thereby of all referring expressions from a standardized language. Both of these consequences are not a priori unacceptable. Indeed, for reasons that mostly had nothing to do with the slingshot, notable authors have gone these ways.⁵¹ However, again it becomes apparent that the immunization of fact causation against the slingshot-argument imposes far-reaching theoretical restraints on the representative of fact causation.

All in all, the two versions of the slingshot-argument discussed here do not conclusively refute fact causation. Still, the slingshot, depending on the back doors chosen, imposes consequential theoretical restrictions upon the friends of fact causation, as, for instance, the invalidity of PSLE and PSST or of \neg -CONV and \neg -SUBS for factual causal statements, the denial of the informal equivalence of (S_1) and (S_2) , or the complete elimination of all referring expressions from a standardized language. These restrictions, to say the least, are inexpedient. For, due to the slingshot-argument, fact causation either cannot reproduce and account for common (coarse-grained) causal judgements, requires the rejection of a theoretical background that has proven valuable in other philosophical contexts, or presupposes extensive additional theoretical work. This is even more unfavorable to the theoretical prospect of fact causation in view of the existence of an analysis of the causal relata that is completely unaffected by slingshots of either type: event causation.

The Argument from Spatiotemporal Locality

Token level causes and effects commonly are taken to be particular dated entities in nature. Spatiotemporal properties apply to events, but not to facts (cf. p. 24). This prima facie shows that events come much closer to what causes and effects are intuitively taken to be. The intuitive plausibility of event causation can be countered by representatives of fact causation in twofold ways: Either (a) facts are analyzed – contrary to the above fact conception – to exist in time and space just like events, or (b) our prima facie intuitions concerning the spatiotemporal characteristics of causes and effects are shown to be a mere guise, perhaps suggested by certain features of causal statements. Authors offering fact conceptions that suit (a) are e.g. Baylis (1948) or Reichenbach (1947). According to Baylis, facts are objects of denotation. They are denoted by true propositions. Facts are spatiotemporally located.⁵² They exist in the same way objects or events exist. Against this background, of course, facts are characterized by all the features commonly attributed to causes and effects. Yet, Baylis' analysis blurs the difference between facts and

⁵¹ We have already seen that Wittgenstein is an example for a skeptic concerning the meaningfulness of the identity predicate (cf. Wittgenstein (1995), §§5.3 et seq; or p. 35 above), whereas Quine, as is well known, has professed the eliminability of all referring expressions from a standardized language (cf. Quine (1982)).

⁵² Similarly Mellor (1995), e.g. p. 9.

events, for facts, thus understood, come very close to being nothing but events. Consequently, Reichenbach, who analyzes facts analogously, explicitly abandons the difference between facts and events. Needless to say that Baylis' and Reichenbach's fact conception is not to be rejected just for blurring the difference between events and facts. However, it renders the controversy over fact vs. event causation merely terminological. Whether causes and effects – understood to be entities with spatiotemporal localities – are called “events” or “facts” is unsubstantial. What philosophically matters is whether causes and effects are attributed spatiotemporal existence or not, and in this respect representatives of fact causation that opt for (a) agree with the event camp.

Arguments against the plausibility of event causation that take route (b) by casting doubt on the spatiotemporal existence of the causal relata usually proceed from the widespread causal interpretation of absences or non-occurrences. Statements like (8) to (11) suggest there being something – whatever it may be called, “event” or “fact” – in time and space that is Brutus' attack on Caesar and something that is Caesar's death. Yet, consider the following causal statements:

- (15) The absence of traffic lights at the intersection caused the crash of two cars.
- (16) The fact that there were no traffic lights at the intersection caused the fact that two cars crashed.

What spatiotemporal entity, according to (15) and (16), caused the crash? Neither of the two statements suggests the spatiotemporal existence of anything causing crashes. On the contrary, both are negative existential statements. They deny the existence of traffic lights at a certain intersection. Still, we are ready to interpret such non-existences as causes. Someone who nonetheless would want to endorse that token level causes are always spatiotemporally located could try to point to the air or the sidewalk where the traffic lights were supposed to be and maintain that it really is the presence of air instead of traffic lights that caused the crash. Obviously, the consequences of such a vindication of the general spatiotemporal localization of causes and effects are absurd. For, neither air nor sidewalks cause accidents. What causes the accident is not the existence of anything instead of traffic lights, but *the absence* of traffic lights or, as the proponent of fact causation would put it, *the fact* that there were *no* traffic lights on site. There does not exist or happen anything causally interpretable near the intersection at the time of the accident. Even though this argument does not demonstrate that no causes and effects are spatiotemporally located, representatives of fact causation pointing to the widespread causal interpretation of non-occurrences, nonetheless, succeed in showing that causal intuitions to the effect that *all* causes and effects are spatiotemporally located are rash. Yet, even if absences of traffic lights are not spatiotemporally located, Brutus' attack on Caesar clearly is. Some causal relata are spatiotemporally located, others are not. Which ones are, which ones are not?

The answer to this question can only be hinted at here. As indicated in section 1.2.1 above, Humean analyses of causation consider general causation (or causal relevance) to be their primary analysandum. In order to account for the relata of

general causation, theories of causation, regardless of their preference for events or facts, ordinarily distinguish between *positive* and *negative causal factors*. The stabbing of Caesar, the crashing of two cars or lightning strokes are examples of positive causal factors, while the absence of traffic lights, the absence of a crash or the absence of lightning strokes constitute negative causal factors. In the present context the term *causal factor* shall be implemented as a neutral term, one that is noncommittal concerning event or fact causation.⁵³ It is a generic term, one that does not designate singular causes or effects on token level, but types of causes or effects. Negative factors, in contrast to their positive counterparts, somehow involve negation. How exactly this is to be understood does not need to be clarified for the purposes at hand. A rough understanding of the difference between positive and negative causal factors, guided by the few examples encountered in this paragraph, will suffice.⁵⁴ With these conceptual distinctions at hand, the question as to what causes are spatiotemporally located and what causes are not, can more specifically be seen as a question about which causal factors have spatiotemporal token causes as instances and which ones do not. The following answer lies at hand now: The instances of positive causal factors are spatiotemporally located token causes, while the instances of negative causal factors are not spatiotemporally located token causes. Brutus' stabbing of Caesar is an instance of a positive causal factor and thus subsists in time and space, while the absence of traffic lights at a certain intersection does not have any token causes in time and space as instances.

The upshot of the arguments from the spatiotemporal locality for the controversy over fact vs. event causation then is a forthright tie. Due to their spatiotemporal locality, events are better suited as instances of positive causal factors, while facts are preferable as instances of negative causal factors.⁵⁵ Clearly though, both event and fact causation have to be further adapted in order to successfully account for all causal relata. Possible adaptations are not hard to find for both conceptions. Friends of event causation could analyze causal dependencies involving absences as cases of general causation and thus as relating factors, not events. Factors then could be spelled out in terms of classes of events, i.e. in terms of abstract, non spatiotemporally located entities. Representatives of fact causation, on the other hand, could try to account for the spatiotemporal locality of instances of positive factors by means of the spatiotemporal locality of fact constituents as objects or coordinates in time and space. It could be argued that intuitions as to the spatiotemporal locality of instances of positive factors actually stem from the spatiotemporal locality of certain constituents of the causal relata, not from the spatiotemporal existence of the relata themselves.

⁵³ For a specification of the notion of a factor cf. section 2.2.2 below.

⁵⁴ For details see Baumgartner and Graßhoff (2004), ch. 2, and section 2.2.2 below.

⁵⁵ In light of this result, it is not surprising that authors such as Bennett (1988) or Mackie (1974) claim that both events and facts can, depending on the context, function as causes and effects.

The Argument from Negation

The fact that instances of negative factors are as readily causally interpreted as instances of their positive complements not only casts doubt on the general spatiotemporal nature of token level causes and effects, but, over and above, serves as grounds for an argument often put forward in favor of fact causation that stems from the possibility to negate facts as opposed to the impossibility to negate events.

Only sentences – the expressions stating facts – can be negated, whereas singular terms – the expressions referring to events – are not subjectable to negation. Proponents of fact causation thus argue that only fact causation provides the formal means to cover all of our common causal judgements.⁵⁶ Is the impossibility to negate singular terms compelling grounds to reject event and favor fact causation? That a sentence is the only linguistic entity open for negation is undisputable. Yet, we have already seen that linguistic entities as such are not suited as relata of causation. It has to be further asked what corresponds to a negative sentence, or rather, what makes a negative sentence true. Representatives of a worldly fact conception as e.g. Russell answer: a negative fact.

If I say ‘There is not a hippopotamus in this room’, it is quite clear there is some way of interpreting that statement according to which there is a corresponding fact, and the fact cannot be merely that every part of this room is filled up with something that is not a hippopotamus. (...) I think you will find that it is simpler to take negative facts as facts, to assume that ‘Socrates is not alive’ is really an objective fact in the same sense in which ‘Socrates is human’ is a fact.⁵⁷

What makes a negative sentence true, according to Russell, thus is not just the given positive facts, but something over and above positive facts. Negative facts, according to Russell, are as objective as positive facts. This point is heavily disputed in the literature.⁵⁸ Why should it be anything above the fact that a table is green that makes the sentence “The table is not red” true? If a positive (worldly) fact is analyzed as a relation holding between localities in time and space, how can a negative fact, i.e. the absence of such a relation, as Russell claims, be just as objective as a positive fact? A (worldly) positive fact – it might be argued – is constituted of entities in time and space, the same cannot be true of a negative fact, for these facts amount to the absence of some relation in time and space. Such things as negative objects, properties, or relations do not have spatiotemporal existence. Or at least there is no need to postulate the spatiotemporal existence of such negative entities for an analysis of causation, because, as we have seen above, the fact that we are ready to causally interpret instances of negative factors shows that a *prima facie* intuition according to which token level causes and effect or their constituents are always spatiotemporally located cannot be upheld against

⁵⁶ Cf. e.g. Mellor (1991 (1987)), Mellor (1995), pp. 131-135.

⁵⁷ Russell (1977), pp. 213-214.

⁵⁸ An overview over the arguments against Russell can be found in Oaklander and Miracchi (1980).

theoretical scrutiny. Intuitions concerning spatiotemporal characteristics of causes and effects have to be adapted. In order to account for instances of negative causal factors, there is no need to advance any spatiotemporally located entities – neither negative events nor negative facts.⁵⁹

The impossibility to negate singular terms and thereby the linguistic representatives of events along with the negatability of fact expressing sentences, hence, is a purely formal argument in favor of fact causation. Fact causation provides the formal means for a straightforward logical analysis of factual causal sentences involving non-occurrences. Any sentence as “ Fa causes Gb ” is unproblematically transferrable into its negative counterpart “ $\neg Fa$ causes $\neg Gb$ ”. Or a factual causal sentence as (16) is readily formalized as “ $\neg Ha$ causes Jb ”. Event causation, on the other hand, analyzes a token level causal dependency in terms of a binary relation as Cab , where C stands for “...causes...”. Therefore, due to formal restraints, a formalization of a sentence as (15) by simply negating the name of the cause is not available. $C\neg ab$ is not a well-formed formula. This constraint, however, does not amount to the impossibility of formalizing causal sentences involving non-occurrences within in the framework of event causation. Other formalizations are feasible. (15) could, for instance, be formalized thus:

$$\forall x(Cxb \rightarrow \neg Hx), \quad (2.14)$$

where H stands for “...is an operating traffic light” and b designates the accident. Relative to a specific theory of causation, yet other formalizations can be thought of. Within the framework of a regularity account, for instance, which essentially analyzes causes as sufficient conditions of their effects,⁶⁰ (15) could also be taken as a statement about type level causation. This might yield a formalization somehow along the following lines:

$$\neg \exists x Hx \rightarrow \exists y Jy, \quad (2.15)$$

where J represents “...is an accident”. No doubt, neither (2.14) nor (2.15) are straightforward and much more will need to be said about the formalization of causal statements involving non-occurrences within an event causation framework.⁶¹ Nonetheless, formalizing sentences as (15) by quantifying over events and, thus, sticking to a pure event ontology is not downright impossible. What formalization of (15) is chosen, in the end, hinges on how the causal relation is analyzed – in terms of regularities, probabilities, or counterfactual dependencies.

The Argument from the Progressive Localization of Causes

What Mackie takes to be the main argument in favor of event causation is our widespread causal ignorance. Fact causation, in his view, calls for naming all of

⁵⁹ Cf. Lewis (2000), pp. 195-196.

⁶⁰ As the study at hand will substantiate, this is highly oversimplified characterization of modern regularity accounts.

⁶¹ Cf. section 3.6.4.

the actually causally relevant features of a cause, whereas event causation does not. On the contrary, the latter offers grounds for what Mackie calls the *progressive localization of a cause*.⁶² For an illustration of what he means therewith, consider the following example: Suppose a police officer is searching for the cause of the explosion that occurred on Elm Street at 2 a.m. on June 2, 2004. Investigating the scene, he finds traces of TNT and the remains of a detonator connected to electrical conductors that lead to a nearby parked car. He now vaguely stipulates that the cause of the explosion must have been

- (17) some occurrence having taken place in that car before 2 a.m. on June 2, 2004, which by means of the conductors set off the detonator.

Upon opening the car he finds that the conductors are connected to a red box with a black button on it. He diagnoses that the cause of the explosion must have been

- (18) some manipulation of the box in the car involving the button before 2 a.m. on June 2, 2004, which by means of the conductors set off the detonator.

He then breaks open the box and finds the electrical conductors connected to a charged battery and a trigger switch. The policeman now specifies the cause of the explosion along the following lines:

- (19) Someone's pushing of the button on the box in the car a tick before 2 a.m. on June 2, 2004, which triggered the switch and by means of the charged battery sent an electrical impulse through the conductors and thereby set off the detonator.

After weeks of further investigation he finds that Tom was the person who pushed the button. He now terminates his causal inquiry by determining the cause of the explosion to be:

- (20) Tom's pushing of the button on the box in the car a tick before 2 a.m. on June 2, 2004, which triggered the switch and by means of the charged battery sent an electrical impulse through the conductors and thereby set off the detonator.

Upon advancing from (17) to (20) the investigated cause is gradually specified. All of these diagnoses refer to one and the same cause of the Elm Street explosion, although on different levels of specification. As the specification becomes more and more fine-grained, the cause and its location are progressively isolated. This form of progressive localization of causes is one of the most commonly implemented and most effective methodologies of causal diagnosis. Causal analyses typically start from very little knowledge about underlying causal structures that only allows for a broad characterization of the cause and its location. Depending on the requests of a particular investigation, causes and their locality then are increasingly specified. In between (17) and (20) many more descriptions of the cause

⁶² Cf. Mackie (1974), pp. 73, 262.

at hand would be possible on intermediate levels of specification. For reasons of brevity, for instance, (20) could be shortened to

(20') Tom's pushing of the button on the box in the car a tick before 2 a.m. on June 2, 2004.

Moreover, the identity of the cause singled out by (17) to (20') would remain unaltered if e.g. (20) were to be further specified to, say,

(21) Tom's pushing of the *black* button on the *red* box in the car a tick before 2 a.m. on June 2, 2004, which triggered the switch and by means of the charged battery sent an electrical impulse through the conductors and thereby set off the detonator.

Even though (21) spells out causally irrelevant features of the cause of the Elm Street explosion, provided that it picks out the same event as (17) to (20'), it nevertheless correctly identifies the cause under investigation. That is all that is required within the framework of event causation.

Only event causation, Mackie holds, can account for this straightforward handling of our causal ignorance, for this unproblematic singling out of causes, and for this form of efficacious causal diagnosis. There are several reasons as to why fact causation cannot serve as grounds for progressive localizations of causes. First, events, due to their locality in space and time, can be progressively located by gradually pinning down their exact location. It makes sense to state of a cause-event that it has taken place somewhere within a certain region of spacetime and then continuously narrowing down that region. An analogous statement involving a cause-fact, in contrast, is nonsensical, because facts are not spatiotemporally located to begin with. Second, as we have seen above, it is irrelevant by means of what singular term an event is referred to, as long as unique reference is guaranteed. The identity of facts – understood in Russellian terms –, on the other hand, crucially rests on the predicates and descriptions used within the corresponding fact stating sentences. As a direct consequence thereof, within the framework of fact causation the methodology of progressively localizing causes, as conducted by advancing from (17) to (21), inevitably yields a counterintuitive multiplication of the causes for each thus investigated effect. Would (17) to (21) be reformulated in terms of factual expressions, each thus generated fact reporting expression would state a different fact – both according to identity criteria for facts along the lines of PSLE and PSST as well as of Leibnizian standards.

The proponent of fact causation might try to evade this consequence by suggesting that a factual causal statement that e.g. corresponds to (17) as

(22) The fact that some occurrence has taken place in the car before 2 a.m. on June 2, 2004, caused the fact that the TNT detonated.

is to be analyzed as a conjunction of many elementary facts, one of which being the actual cause of the explosion.⁶³ Then, it could be argued that progressively

⁶³ Cf. p. 27 above.

locating causes is to be understood as progressively minimalizing the conjunction of elementary facts recorded by unspecified statements as (22). This strategy, however, would presuppose standards of reducing colloquial factual causal statements as (22) to a standardized language of elementary facts. Before such reduction standards are available it is hard to evaluate the merits of this proposal.

Yet, apart from a possible analyzability of (22) in terms of conjunctions of elementary statements, there is another reason why fact causation does not offer grounds on which to account for the progressive localization of causes. True factual causal statements seem to presuppose some ‘explanatory connection’ between the causing and the caused fact. In order to illustrate what is meant by ‘explanatory connection’, consider the following example taken from Hart and Honoré (1985 (1959)):

(...) owners of a vessel, in breach of a Hong Kong ordinance, sent a ship to sea without duly certificated officers. The master [named Sinon], though uncertificated, was a perfectly competent man of long experience, but the ship was in fact involved in a collision when he was the officer of the watch and guilty of negligent navigation.⁶⁴

Against the theoretical background of event causation, the cause of the collision can be referred to on different levels of specification and by implementing different emphases:

- (23) Sinon’s navigation before the collision
- (24) The perfectly competent watch officer’s navigation before the collision
- (25) Sinon’s negligent navigation before the collision.

All of these noun phrases designate the same event, some of them explicitly mention features of the referred to event that were in fact causally relevant to the collision, others do not. Yet, whether the cause-event at hand is designated with respect to its causally relevant properties or not, is of no importance to its individuation. The event named in (24) has the property of being a negligent navigation just as well as the event named in (25), which explicitly speaks of negligent navigation.

In contrast to these multifarious designation options made available by event causation, fact causation seems to be constrained to one specific way of stating the causing fact. Analogously to (23), (24), and (25) the fact that the ships collided could, in factual terms, be said to be caused by:

- (26) The fact that Sinon navigated before the collision
- (27) The fact that the perfectly competent watch officer navigated before the collision
- (28) The fact that Sinon navigated negligently before the collision.

⁶⁴ Hart and Honoré (1985 (1959)), p. 119.

While it is at least dubious whether (26) correctly states the cause of the collision, (27) certainly cannot be said to do so. The fact that the perfectly competent watch officer navigated did not cause the fact that the ship got involved in a collision. In (27) the crucial ‘explanatory connection’ between causing and caused facts is missing. Eventually, (28) turns out to be the only expression unquestionably reporting the causing fact of the collision. Without further dwelling on the admittedly fuzzy notion of ‘explanatory connection’ it seems clear that not all of the expressions (26) to (28) can be said to state the cause of the fact that Sinon’s ship was involved in a collision.

Again, the reason why event causation allows for all the different descriptions in (23) to (25) of the cause of the collision stems from event causation only requiring unique referencability to causes. The same cannot be said of fact causation. Taking facts to be the relata of causation asks for some ‘explanatory connection’ between causing and caused facts – how ever the notion of ‘explanatory connection’ might eventually be spelled out.

Summing Up

Most importantly, the discussion of the main arguments in the debate over fact vs. event causation has revealed some crucial requirements that a successful analysis of the causal relata has to fulfill. First, we have seen that common causal intuitions are coarse-grained with respect to the individuation of causes and effects. Causal dependencies are independent of the way they are referred to or described. Causes and effects cannot be proliferated by redescription. This suggests that causation, whether it is analyzed to be a binary relation or a sentential connective, must allow for substitutions *s.v.* of co-referring singular terms and, if it is taken to be a sentential connective, logically equivalent expressions, which, in turn, gives rise to the slingshot-argument. Second, many causes and effects, yet not all of them, are spatiotemporally located. Non-existences or absences can also be causally interpreted. That means that causal relata have to be open for negation. Positive causal factors have instances with a specific location in spacetime, negative factors do not. Third, we repeatedly touched upon the fact that there are at least two different kinds of causal relations, one on token and one on type level. Accordingly, there are at least two kinds of causes and effects: type causes/effects and token causes/effects. This difference will be further elaborated in sections coming up. And fourthly, an analysis of the causal relata should allow for the progressive localization of causes. Causes and effects have to be analyzable on variant levels of specification.

The previous paragraphs have shown that none of the two most widely received analyses of the causal relata is fully satisfactory. For instance, both event and fact causation lack an unproblematic identity criterion for their respective key ontological category – a criterion that fits the needs of a subsequent analysis of causation itself. Moreover, event causation cannot straightforwardly account for instances of negative causal factors, while fact causation faces difficulties when it comes to accounting for the instances of positive causal factors. Causal statements involving

non-occurrences are not easily formalized within an event causation framework, for singular terms are not negatable. Finally, in view of the slingshot-argument, fact causation is either forced to accept an overly fine-grained individuation of causes and effects or to reject a theoretical background that has proven valuable in respects that are unconnected to causation.

We have seen that both competing analyses of the causal relata can be adapted such that they successfully meet the objections raised against them. Nonetheless, the theoretical weightiness of the modifications necessary to immunize the two accounts differ significantly. Although fact causation might be thus strengthened that it evades the consequences of the slingshot, successfully resolves the identity of the causes mentioned in, say, (12) and (13), and even allows for causal analyses on different levels of specification, an amendment of fact causation to these requirements is, if at all, possible only with a considerable additional theoretical effort: Standard inference rules of first-order logic, a referential reading of definite descriptions, or even the meaningfulness of the identity predicate have to be rejected or new standards for the formalization of natural language have to be developed. Event causation, on the other hand, is not affected by the slingshot-argument. Furthermore, if events are analyzed in terms of spatiotemporally localized particulars, event causation accounts for common coarse-grained causal judgments and allows for causal analyses on variant levels of specification. Clearly though, additional work needs to be done as regards the identity criteria of events and as regards the formalization of causal statements involving non-occurrences. These adjustments, however, at most affect the analysis of causation. Event causation offers grounds to meet all requirements of a successful analysis of the causal relata without unclear broad philosophical follow-ups. Therefore, the study at hand will analyze causes and effects within an event causation framework. Details and remaining questions of this framework shall be addressed in the upcoming sections.

2.2.2 *Singular Events vs. Event Types*

Even though events are standardly taken to be the relata of singular causation, as the previous section has shown, the prevalent criteria for event identity provided in the literature cannot unproblematically be implemented within an analysis of causation. Davidson's (1969) criterion, when taken as grounds for a theory of causation, launches an infinite regress, an analysis of causation based on Quine's (1985) account is thus coarse-grained that it induces the possibility of self-causation, and Kim's (1976) criterion yields a counterintuitively fine-grained individuation of causes and effects.⁶⁵ There are many other accounts of event identity available, some involving modal operators as Brand's (1977), others, as Lombard's (1979) or Cleland's (1991), analyzing event identity with recourse to other unclear and, above all, causally entrenched ontological categories as changes or states of

⁶⁵ Cf. p. 25 above.

affairs. None of these accounts either have been or, as I contend, can be satisfactorily used as a fundament of a Humean analysis of causation. They all suffer from one of the following defects: (i) They are overly coarse-grained or (ii) overly fine-grained or (iii) they plainly presuppose clarity on the cause-effect relation. The first desideratum of our analysis of the causal relata in terms of events, therefore, is a criterion of event identity that fits the needs of the subsequently proposed theory of causation. This goal is attained if our criterion is thus fine-grained that it never identifies causes and effects, i.e. that it rules out self-causation, if it is thus coarse-grained that it does not allow for multiplying causes and effects by merely redescribing them, and if it does not presuppose clarity on causation itself. That means, an event notion suitable for our purposes has to avoid defects (i) to (iii).⁶⁶

The most influential among the above event theories, no doubt, is Kim's. Several attempts have been made to stick to Kim's basic analytical strategy of equating events with triples consisting of objects, properties and times, while, at the same time, rendering Kim's criterion of event identity more coarse-grained. The most promising among these attempts are Rosenberg (1974), Bennett (1988) and Bennett (1996). By drawing on these studies I shall thus in the following sketch an event theory that avoids defects (i) to (iii) and then introduce the notion of an event type or *factor*.

Singular Events

Events, as already mentioned, shall be taken to be designatable by singular terms. They are possible values of bound variables, just like physical objects. They are particulars. Again in accordance with physical objects, events are located in space and time. They are token or singular entities, inasmuch as they have exactly one spatiotemporal location. Events do not occur twice. In order to distinguish them from the yet to be introduced types of events, they are, accordingly, often labelled *singular events*. Singular events can be related in terms of *singular causation*. As we have seen in the previous section, singular causation is irreflexive and the subsistence of token level causal dependencies among events is independent of the singular terms or definite descriptions by means of which these events are referred to. Hence, on the one hand a notion of event identity that suits the requirements of an analysis of singular causation must prohibit that events may be identified to cause themselves and, on the other, it must render causal dependencies

⁶⁶ Of course, only a very selective and unrepresentative list of event analyses is provided here. The event literature is vast. A more representative overview can be found in Casati and Varzi (1996). One recent influential further study on events deserves separate mention here: Kanzian (2001). Kanzian analyzes event identity by drawing on event delimiting states of affairs (cf. p. 250). Kanzian's analysis – if used as ontological fundament of singular causation – would in fact not give rise to the difficulties (i) to (iii). Still, it induces another problem: In order to keep the ontology of causation as simple as possible it is common usage to employ a *broad* notion of event that embraces both dynamic events and static states of affairs. Such a broad event notion, clearly, is not clarified along the lines of Kanzian.

on token level independent of the way causes and effects are referred to. Furthermore, any analysis of the event notion that purports to be integratable into a theory of singular causation needs to be broad to the extent that it classifies both static states of affairs as well as state changes as events.⁶⁷ Commonly we are not only ready to causally interpret changes or processes, but also static states of affairs:

(29) The red coloring of this light bulb causes the red shimmering of this room.

(30) The density of this stone caused the breaking of this window.

Colloquially, the notion of event is applied mainly to processes involving some change of states of affairs. Yet, neither red colorings nor densities are events in this restricted sense. Nonetheless, both (29) and (30) are perfectly acceptable causal statements. Professing a wide notion of event – one that encompasses red colorings just as well as lightening strokes – serves conceptual parsimony for the consecutive analysis of causation, such that, whenever a spatiotemporal entity can be causally interpreted, it is susceptible to the label “event”.

Singular events shall be symbolized by italicized small letters a , b , e , e_1 , e_2 etc.⁶⁸ Correspondingly, x , y , z , x_1 , x_2 etc. shall be used as variables running over the domain of events. There is a function τ that assigns exactly one point or interval in time to every event e , e.g. $\tau(e) = t_1$; where t_1 , for instance, is March 15, 44 B.C. Similarly, exactly one locality in space is assigned to each event by a function σ , e.g. $\sigma(e) = s_1$; s_1 , say, being the Forum Romanum. τ assigns a value to every e relative to some standard event, whose temporal location is being presupposed – possible candidates are prototypical radioactive decays or the working of a certain chronometer. σ 's attribution of a spatial locality to every e , on the other hand, is relative to the origin of some coordinate system. Of course, τ and σ are also relativized to a respective scale. Depending on the level of specification chosen for a corresponding event analysis, τ assigns days, hours, or microseconds to events, whereas σ localizes events within cities, houses, square meters, or attributes a single point to an event.

There shall be no pretension to high precision as regards spatiotemporal localizability of events. Spatiotemporal boundaries of events are vague, just like spatiotemporal boundaries of physical objects as mountains or desks.⁶⁹ However, on par with Quine (1985), I contend that this vagueness does not normally affect the discrimination of events, especially of macroscopic events, which will be of primary interest to the investigation at hand. A rough and approximate determination of the spatiotemporal locality of events is mostly sufficient to keep events apart. Thus, since uncertainties with respect to the spatiotemporal location of events only affect a minority of event pairs, this problem will be disregarded in the present context.

⁶⁷ This broad event notion can be seen as common ground on which virtually all theories of causation rest that opt for events as causal relata (cf. e.g. Collins, Hall, and Paul (2004), ch. 1, §7).

⁶⁸ The notational conventions used within the present enquiry are summed up and listed in the appendix, see pp. 261 et seq.

⁶⁹ Cf. Quine (1985), pp. 167-168.

This yields two (uncontroversial) necessary conditions for the identity of events e_1 and e_2 . If e_1 and e_2 do not occur at the same spatiotemporal locality or within the same spacetime zone⁷⁰, they are not identical:

$$\tau(e_1) \neq \tau(e_2) \rightarrow e_1 \neq e_2 \quad (2.16)$$

$$\sigma(e_1) \neq \sigma(e_2) \rightarrow e_1 \neq e_2. \quad (2.17)$$

As we have seen in the previous section (p. 25), $\tau(e_1) = \tau(e_2) \wedge \sigma(e_1) = \sigma(e_2)$ is not sufficient for the identity of e_1 and e_2 . Davidson's metal ball that rotates and heats up at the same time illustrates that there are different events that occur at identical locations in space and time. (2.16) and (2.17) need to be complemented and strengthened in order to obtain a sufficient condition of event identity.

A closer look at our linguistic reference to events shall turn out to be informative with respect to a possible supplementation of (2.16) and (2.17). In colloquial language reference to events is normally made by means of definite descriptions. The singularity of events is expressible by indexicals as in (31) or by explicit relativization of the descriptions to a spatiotemporal location as in (32) or (33).

(31) This explosion

(32) The lethal attack on Caesar on March 15, 44 B.C., in the Forum Romanum

(33) The 2004 Summer Olympics in Athens.

Of course, except for definite descriptions that have come to function as proper names as in (33), colloquial talk regularly drops explicit mention of spatiotemporal coordinates, whenever the context clarifies reference. Furthermore, certain event descriptions guarantee singularity by means of the unique features of the events described, e.g.:

(34) The assassination of John F. Kennedy

(35) The first human step on the moon

(36) The sinking of the Titanic.

(33) and (36) illustrate that events can be of high complexity, i.e. events may well be analyzable in terms of simpler events. The 2004 Summer Olympics in Athens may be said to consist of the aquatics and the archery and ... and the wrestling competitions between August 13 and 29, 2004, in Athens. The present context does not require preference for a specific level of complexity or specification. On the contrary, the theory of causation to be developed subsequently will explicitly leave room for causal analyses on different levels of specification. Moreover, it will allow for shifts of specification. It therefore does not presuppose causal analyses to be conducted with respect to events which by some arbitrary criterion are marked to be atomic.

⁷⁰ Cf. Bennett (1988), p. 12.

While, as Davidson (1980 (1967)) or Bennett (1996) emphasize, the metaphysics of events must not be confused with considerations concerning our linguistic reference to events, the latter nonetheless reveals something important about the former. All of the definite descriptions (31) to (36) have one thing in common: They all pick out an event by virtue of a property exemplified by an object or instantiated within a spacetime zone. Thus, Bennett (1996) is right in claiming that Kim (1976) correctly analyzes the metaphysics of events: Events – in the broad sense – constitute a composite ontological category that is made up of properties, objects and times or, more generally, spacetime zones. However, as Bennett (1996) again correctly indicates, Kim gets the semantics of event descriptions wrong. According to Kim, event descriptions pick out an event by drawing on its one and only constituting property, by means of which events are then identifiable. Take, for instance example (34) above. The assassination of John F. Kennedy, which, in a Kimian vein, is analyzed in terms of the triple $\langle a, P, t_1 \rangle$, where a stands for Kennedy, P for the property of being an assassination and t_1 for November 22, 1963. Kim holds that any other definite description does not refer to Kennedy's assassination unless it picks out the tragic event in Dallas by drawing on its constitutive property of being an assassination. This semantics of event descriptions lies at the heart of Kim's identity criterion for events, which, as we have seen in the previous section, yields an overly fine-grained ontology of causes and effects.

Even though an event may be successfully picked out linguistically by indicating a single property exemplified at a spacetime zone, events ordinarily feature a host of properties. Consider anew Brutus' lethal attack on Caesar on March 15, 44 B.C., in the Forum Romanum and take a to refer to this event. a is not only a lethal attack as indicated in (32), but moreover:

- (37) a stabbing
- (38) an infliction of a wound
- (39) a killing of an emperor
- (40) an action of a commanding officer
- (41) a jerky movement of a human body

Events can be referred to on various levels of specification, by means of descriptions involving various predicates and names. Reference to events is not necessarily affected by substituting one definite description for another, such that one mentions the killing of an emperor and the other the lethal attack on Caesar. As long as reference to spatiotemporal localities remains unchanged, i.e. as long as the antecedents of (2.16) and (2.17) are not satisfied, the designated events may very well be the same.

However, if definite descriptions referring to a are substituted by definite descriptions referring to an event b that does not share all properties with a , a and b – even if b spatiotemporally coincides with a – are not identical. Take b , for instance, to refer to Brutus' shouting at Caesar during his attack. a and b occupy – at least

relative to the level of specification implemented in this example – the same space-time zone, both a and b feature (40) – and possibly even (41). Yet, not both of these events are characterized by properties (37) to (39). a is a stabbing, which b is not, and b is a shouting, which a is not. Moreover, a is a cause of Caesar’s death, while b is not. Or take an event c that consists in Brutus’ heart beating at over 150 beats per minute during the attack on Caesar. c spatiotemporally coincides with a even relative to highly specific spatiotemporal scales. Nonetheless, c differs from a , for it is not characterized by the properties (37) to (41). Again, a is a cause of Caesar’s death and c is not. Finally, the reason why the rotation and the simultaneous heating up of Davidson’s metal ball⁷¹ differ is at hand now: Even though the two spatiotemporally coinciding events both are physical processes or accelerations of metal molecules, they do not share all properties – one is a rotation of a metal ball, the other a heating up of a metal ball.

Now, compare event a referred to above and event d : Brutus’ stabbing of Caesar on March 15, 44 B.C., in the Forum Romanum. As a , d is a jerky movement of a human body, an action of a commanding officer, a lethal attack, or an infliction of wound. a and d not only spatiotemporally coincide, but moreover share all properties. All of a ’s properties are featured by d as well and vice versa. This finding can be generalized: Whenever a singular term refers to an event e_1 , that spatiotemporally coincides with e_2 and that, in addition, has all the properties e_2 has and e_2 has all the properties e_1 has, e_1 is identical to e_2 or, in other words, the referent of the first singular term is identical to the referent of the second.

(2.16) and (2.17) can now be complemented to render a both necessary and sufficient condition for event identity that does not give rise to the problems (i), (ii) and (iii) when used to individuate singular causes and effects. Events e_1 and e_2 are identical – or, if preferred, singular terms referring to e_1 and e_2 co-refer – iff they have identical spatiotemporal locations and for all properties Z , e_1 has Z iff e_2 has Z .⁷² Or symbolically:

$$e_1 = e_2 =_{df} \tau(e_1) = \tau(e_2) \wedge \sigma(e_1) = \sigma(e_2) \wedge \forall Z(Ze_1 \equiv Ze_2). \quad (EI)$$

Emphasis needs to be placed on the fact that sharing all properties is independent of the level of specification on which definite descriptions refer to events e_1 and e_2 . Assume e_1 to be referred to by (42) and e_2 to be the event designated by (43).

(42) Tom’s drinking of a beer at noon on September 7, 2004

(43) Tom’s having a drink made of hops, malt, and water containing 4.4% of C_2H_6O by volume at noon on September 7, 2004.

Even though (43) is by far more specific than (42), both descriptions refer to coinciding events that share all properties. e_1 is a drinking of a beverage made of hops, malt, and water just as e_2 is a drinking of a beer. Both e_1 and e_2 are alcohol

⁷¹ Cf. p. 25 above.

⁷² For a similar proposal cf. Rosenberg (1974).

consumptions, both are human actions involving a bottle, and both are intakes of C_2H_6O etc. (42) and (43) thus refer to the same event. e_1 is identical to e_2 .

Note that (EI) is not claimed to adequately analyze event identity for any theoretical context in which the ontological category of events is of interest. During the past 50 years or so, event ontology has become a self-contained philosophical branch, largely separated from ensuing theories building on the event fundament. This disciplinary separation has proven particularly disadvantageous to the philosophy of causation, as it has led to the sheer lack of a criterion for event identity that could be satisfactorily implemented within an analysis of causation. Available accounts of event identity suffer from one of the defects (i) to (iii). Instead of attempting to capture event identity in a way that suits the needs of all philosophical areas concerned with events, I here content myself with tailoring event identity to the demands of an analysis of causation, more specifically, to the demands of a theory of causation that subscribes to the Humean tradition. (EI) meets these demands as it avoids defects (i) to (iii).

Clearly, (EI)'s avoidance of (i) to (iii) comes at the price of its presupposition of the notions of a property and of a property exemplification or of having a property. As is well known, the ontology of properties is no less intricate and controversial than the ontology of events. What is the ontological make-up of properties? What does it mean for an event to have a property or for an object to exemplify a property? Or linguistically put: Under what conditions does an event e satisfy a predicate Z ? What are the truth-makers of statements of type Ze ? These are all difficult questions – especially for the Humean – which, however, cannot be addressed here. In order to guarantee the applicability of (EI), it shall simply be presupposed that for every event x and for every predicate χ it is decidable whether χx or $\neg\chi x$.

Event Types, Factors

We have seen on several occasions now that there are two different causal relations, one on token and one on type level.

(44) Shamus' drinking of 6 beers at noon on September 7, 2004, caused Shamus' drunkenness in the afternoon of September 7, 2004.

(45) Drinking alcohol causes drunkenness.

In (44) "...cause/d..." relates singular events, whereas the relata of "...cause/d..." in (45) are not uniquely spatiotemporally located entities as events. Instead, (45) relates types of events, *event types* or *factors* for short. The predicate "...cause/d..." has a different domain in (44) and in (45). As the relata of "...cause/d..." differ, this predicate is equivocal in (44) and (45). It represents different relations in both sentences. (44) is a sentence about singular causation, while (45) expresses general causation. In order to clearly distinguish the two kinds of causal relations, the predicate "...cause/d..." shall be reserved for statements

about singular causation from now on. For general causal statements a new predicate is introduced: *causal relevance*. For reasons of conceptual simplicity, however, the terms *cause* and *effect* will be applied both to events and factors. Given these terminological conventions, (44) is rephrasable as (44'), while (45) must be rephrased in terms of (45') or (45'').

(44') Shamus' drinking of 6 beers at noon on September 7, 2004, is the/a cause of Shamus' drunkenness in the afternoon of September 7, 2004.

(45') Drinking alcohol is causally relevant to drunkenness.

(45'') Drinking alcohol is a cause of drunkenness.

Section 2.3 below will be concerned with causal relevance and the important differences between singular and general causation. This paragraph introduces the relata of general causation: event types or factors.

Notwithstanding the singularity of events, among different singular events a great many similarities hold. Tom's reading of the newspaper today shares countless features with Tom's reading of the newspaper yesterday, even though both events have different spatiotemporal localities. The two event tokens are of the same type, they are both newspaper readings by Tom. Every day many events occur that are type identical to events that occurred yesterday. Every day the sun rises, the tide goes out and comes in, a car engine is started and stopped innumerable times throughout its life, and rain is falling over and over again. All of these type identical singular events form similarity sets, e.g. the set of rainfalls or the set of engine starts etc. Event types shall be analyzed as such similarity sets. They are sets of type identical singular events, of events that share at least one feature. Contrary to singular events, event types are generic entities. They are not located in time and space themselves, but they are *instantiated* in time and space by singular events. Whenever a member of a similarity set that corresponds to an event type occurs, the latter is said to be instantiated. Examples of event types are:

(46) Rainfall

(47) Breaking of a window

(48) Red coloring of a light bulb.

(48) shows that, due to the wide event notion introduced in the previous paragraph, event types are not only sets of changes or processes, but may also be sets of (static) states of affairs. In order to reduce irritation of common linguistic intuitions as to the non-static character of events, in case of event types recourse is often made to the neutral term "factor" which, as mentioned above, shall be used synonymously with "event type".

As singular terms designate singular events, predicates can informally⁷³ be seen to stand for factors. The extension of these predicates then corresponds to the

⁷³ Below a set theoretic representation of factors will be introduced. However, for easier readability and for reasons of brevity factors will subsequently mostly be represented by expressions as in (46) to (48).

similarity set that constitutes the respective factor. Thus, over and above representing factors, predicates *define* factors. For instance, the factor mentioned in (46) is definable as the set of events which satisfy "... is a rainfall". However, not any predicate represents or defines a factor. There are several important restrictions as to what predicates can be used in factor definitions. The notion of a factor is introduced with respect to the analysis of causation. Therefore, whatever is to count as a factor must be serviceable to that analysis and implementable in causal enquiries. Any predicate that forecloses such an implementation shall not be accepted in factor definitions. On these grounds, predicates of the following sorts shall *not* be taken to define factors. Each restriction will be clarified subsequently.

- (a) predicates that apply to *less than 2* events
- (b) predicates that apply to *all* events
- (c) *local* predicates
- (d) predicates that involve *causal* notions

The reason for restriction (a) stems from the fact that causal judgments, both on token and on type level, ask for empirical justification – they must be testable, at least in principle.⁷⁴ Causal judgments, however, that involve uninstantiated event types, evidently, cannot be empirically warranted. Testability presupposes the existence of instances of causally analyzed factors. Predicates with empty extensions in the domain of events cannot meaningfully be seen as representing something that is causally interpretable. Therefore, predicates that do not apply to any event do not define or stand for factors.⁷⁵ Moreover, as will be shown in section 3.6.2, causal structures are of a minimal complexity that involves at least two alternative causes of a given effect. These alternative causes, in turn, are required to demonstrate their causal relevance independently of each other in at least two separate test situations.⁷⁶ In each of these test situations the corresponding effect must be instantiated for the purported causal relevance to be established. Hence, factors can only be interpreted as effects if they have *at least two* instances. Since no factor shall be excluded from such an interpretation by mere definition, predicates with event extensions including exactly one element shall not be admitted in factor definitions. Every factor defining predicate must apply to at least two events. Of course, in experimental practice that builds on arbitrary reproducibility of test situations, factors with exactly – or little over – two instances will hardly be of any use or interest. Depending on the significance criteria implemented in concrete causal investigations, it will be reasonably clear what is to count as an acceptable amount of instances for each factor and what is not. Based on such pragmatic

⁷⁴ The question as to how causal hypotheses are empirically tested will be addressed in all detail in chapter 5. For now, a rough understanding of causal reasoning along the lines of Mill's Method of Difference suffices (cf. Mill (1879 (1843)), p. 256).

⁷⁵ Cf. section 1.2.4.

⁷⁶ Cf. section 2.3.5.

considerations the set of admissible factor defining predicates might easily be further narrowed down. (a) just indicates the absolute minimum amount of instances for a factor, below which the latter would not be conclusively causally analyzable *by definition*. Furthermore, even in the optimal case a factor with exactly two instances can only be integrated in the aforementioned minimal causal structure that involves exactly two alternative causes. As soon as the place of a factor within a more complex structure is investigated, the number of instances needed increases significantly.⁷⁷

As shown shortly, factors shall be taken to be negatable. Since negative factors must satisfy the same conditions with respect to testability of their causal relevancies, predicates embracing all events are not acceptable for factor definitions, for negations of all-embracing predicates are empty. Restriction (b) will be accounted for in more detail below (p. 63).

Restriction (c) involves the notion of locality of predicates, which, as is well known, has been discussed extensively in the literature.⁷⁸ (c) excludes predicates that, even if they satisfy (a) and (b), involve spatiotemporal constraints as, for instance, "...is a New Year's Eve fireworks between 1980 and 1990" or "...is a New Year's Eve fireworks in Shamus' garden". The reason for excluding these so-called *local* predicates in the context at hand does not lie, as normally in studies on laws of nature, in metaphysical considerations to the effect that – due to conceptual presumptions – causal regularities or laws are not taken to regulate local domains only. Rather, local predicates shall be excluded for they would define factors that are not arbitrarily reproducible *by definition*. It follows from the mere definition of locally defined factors that they do not have instances beyond their local domain. For instance, there will never again be a New Year's Eve fireworks between 1980 and 1990, or outside of Shamus' garden there are no New Year's Eve fireworks in Shamus' garden. Arbitrary reproducibility will prove to be of crucial importance to causal reasoning. Of course, many non-local factors are not arbitrarily reproducible either, due to e.g. limited financial resources or due to the fact that their instances lie outside the domain of human intervention. Nonetheless, repeated instantiations of such factors are not excluded by definition.

Armstrong (1983) claims that every predicate involving local constraints is replaceable by a co-extensional non-local predicate.⁷⁹ For instance, "...is a New Year's Eve fireworks between 1980 and 1990" could be replaced by a disjunction of arbitrary non-local properties that each happen to apply to exactly one of the many (but finite) New Year's Eve fireworks between 1980 and 1990. It should not be difficult to find such unique properties for each of these fireworks, e.g. the specific combination of pyrotechnic devices used or the exact amount of people

⁷⁷ Chapter 5 will spell out how many instances are required in order to locate a factor within a causal structure of a given complexity.

⁷⁸ Cf. e.g. Popper (1994 (1934)), ch. 5, Hempel (1977 (1965)), ch. 2, Goodman (1983), or Armstrong (1983).

⁷⁹ Cf. Armstrong (1983), ch. 1. See also section 1.2.7 above.

watching etc. Still, while the local predicate is not admissible for factor definitions, a disjunction of non-local predicates might well be. For the latter does not exclude repeated instantiations of the corresponding factor by definition. Referring to the New Year's Eve fireworks between 1980 and 1990 by enumerating characterizing properties of each of these fireworks does not foreclose the reproducibility of such fireworks on definitional grounds. It might well be that there will be another New Year's Eve fireworks in the future that has the same characterizing properties as one of the fireworks between 1980 and 1990. Even if that is not the case, predicates specifying the characterizing properties of fireworks between 1980 and 1990 are acceptable in factor definitions, because the non-recurrence of events satisfying those predicates will most likely be due to lack of causal interest in factors of this kind. The instances of such factors do not recur because nobody is interested enough in their causes to set up costly and time-consuming causal tests that would reproduce their instances. Nonetheless, such reproductions would be possible, at least in principle.

It must be stressed though that excluding local predicates from factor definitions does by no means block inquiries into the causes of, say, the 1984 New Year's Eve fireworks in Bern. This event instantiates a vast amount of perfectly well-defined factors, as being a fireworks or a pyrotechnic demonstration or a waste of tax money etc. Each of these factors, given adequate evidence, can be integrated in causal structures of any chosen complexity. Once a causal structure involving factors that are instantiated by the 1984 New Year's Eve fireworks in Bern is available, conclusions with respect to the causes of the event under investigation can be drawn. In a similar vein, analyses of the causes of e.g. the breakout of World War One or the first human step onto the lunar surface might be conducted, provided that no predicates as "... is the breakout of World War One" or "... is the first human step onto the lunar surface" – both violating (a) – are used in defining the analyzed factors. Events instantiate a vast amount of factors. Investigations into the causes of events are not constrained to one specific factor a given event instantiates. Rather, depending on the specific interest of a causal enquiry, the analyzed event must be taken as instance of different factors.

Finally, restriction (d) prohibits conceptual circularity. Only non-causally defined factors are serviceable to causal investigations. Thus, predicates as "... is the cause of a fire", although furnished with event extensions that satisfy (a), (b), and (c), would, if accepted in the definiens of factors, yield factors that are of no use to causal enquiries. Therefore, they are excluded from factor definitions.

All in all, predicates as the following do not define factors:

- ... is a table
- ... is the sinking of the Titanic
- ... is spatiotemporally located
- ... is an event
- ... occurs
- ... does not occur

- ... causes an explosion.

Factors are symbolized by italicized capital letters A, B, E, E_1, E_2 etc., with variables Z, Z_1, Z_2 etc. running over the domain of factors.⁸⁰ In order to clearly distinguish factors from predicates, the latter shall exclusively be symbolized by non-italicized capitals A, B , etc. An event type as “running car engine” (A) can be defined as the set $\{x : x \text{ is a running of a car engine}\}$, or formally:

$$A =_{df} \{x : Ax\}, \text{ such that } A \text{ satisfies (a), (b), (c), and (d).} \quad (2.18)$$

“ A ” is interpretable as any predicate satisfied by a singular event x iff x is a running of a car engine. “ A ” could thus be interpreted as either one of the following predicates:

- ... is a running of a car engine
- ... is an operating car engine
- ... is a running of an Alfa Romeo engine or of an Audi engine or ... or of a Volvo engine
- ... is a traveling down of a car engine piston on an intake stroke, up on a compression stroke, down on a power stroke, up on an exhaust stroke, and back down on an intake stroke etc.

Assuming that all these predicates are co-extensional, they can all be used to define the factor “running car engine”. In light of the criterion for event identity presented in the previous section, factor identity can now be straightforwardly accounted for in terms of set identity. For a factor $A =_{df} \{x : Ax\}$ and a factor $B =_{df} \{x : Bx\}$,

$$A = B =_{df} \{x : Ax\} = \{x : Bx\}. \quad (FI)$$

Alternatively factor identity can be spelled out contextually:

$$A = B =_{df} \forall x(Ax \equiv Bx). \quad (FI')$$

Or in words: A and B are identical iff the predicates in the definiens of A and the predicates in the definiens of B are co-extensional.

Factor identity is thus analyzed on purely extensional grounds. This might raise similar objections as have traditionally been raised against e.g. Quine’s reduction of predicate meaning to predicate extension or against his analysis of properties in terms of classes of objects having the corresponding property.⁸¹ As is well known, Carnap (1947), for instance, points to co-extensional predicates as “... is a creature with a kidney” and “... is a creature with a heart” that are not synonymous, therefore, predicate meanings and extensions cannot be identified. Similarly for the identification of classes of entities having a property and the respective property:

⁸⁰ Letters X and Y are not taken to be factor variables for they will be given a special function in chapter 3.

⁸¹ Cf. Quine (1960), §§12, 25, 43, Quine (1980a), Quine (1980c).

The class of creatures with kidneys and the class of creatures with hearts may be identical without the two properties of having a kidney and of having a heart being identical as well. Building on this debate, an analogous objection to (FI) could be raised. It might be argued that, although the factors C and D listed below are defined by means of co-extensional predicates, there are good reasons not to identify C and D in causal contexts.

$$C =_{df} \{x : x \text{ is an arrival of a creature with a kidney}\}$$

$$D =_{df} \{x : x \text{ is an arrival of a creature with a heart}\}$$

The non-identity of C and D might be argumentatively backed by indicating a causal processes in which one of the two factors is claimed to be involved, but not the other. One possible scenario could be the following: Tom's hated boss enters a hospital room in which Tom is lying, suffering from a serious heart disease. Tom is waiting for a heart transplantation. Upon the arrival of his boss, he gathers all his remaining forces and strikes dead his visitor with the bedside lamp in order to generate a heart donor. It might be said that the boss' arrival being an instance of D is causally relevant for the lethal attack, while the same does not hold for C .

What could be the argumentative basis for such differences as regards the attribution of causal relevance to C and D ? The only basis that justifies causal judgments, as mentioned above, is empirical to the effect that differences in the attribution of causal relevance must be backed by differences in the empirical behavior of the corresponding factors. This requires at least one instance of D to be involved in a causal process in which no instance of C is involved or vice versa, which, however, given definitions of C and D in terms of co-extensional predicates, is not the case for C and D . Whenever and wherever C is instantiated, D is instantiated as well and vice versa. Hence, there is no test situation that empirically justifies differences in the attribution of causal relevance to C and D . Furthermore, the above scenario might easily be reanalyzed such that causal intuitions as to the seeming differences with respect to the causal relevancies of C and D become empirically warrantable. Take factor C' to be the knowledge about the arrival of a creature with a kidney and factor D' to be the knowledge about the arrival of a creature with a heart. C' and D' are not defined with recourse to co-referring predicates and, thus, differences in the causal relevancies of C' and D' might, given appropriate evidence, easily be empirically validated.

At this point, the opponent of (FI) will resort to counterfactual conditionals claiming that, had it not been the case that D was instantiated before Tom's strike, the latter would not have happened, whereas the same cannot be said of C . Had it not been the case that C was instantiated before Tom's strike, the latter might nonetheless have happened – if D had been instantiated. The opponent of (FI) here alludes to a hypothetical test situation in which D but not C is instantiated and which, therefore, could serve as licensing grounds on which to attribute different causal relevancies to C and D . Either such a test situation is empirically realized, in which case C and D would no longer be defined by means of co-extensional

predicates, or such a test situation is not realized, in which case there is no empirical basis for differences in the causal interpretation of C and D .

Of course, much more as regards the criteria of empirical testability must and will be said as we proceed. For now, all that matters is the following: Whatever one's position might be concerning the semantic debate about the identifiability of predicate extensions and predicate meanings, all that is of interest for causal contexts are differences in the empirical behavior of factors. Among factors there can only be empirical differences if one of two factors is instantiated by events that do not at the same time instantiate the other factor. Given that causal reasoning exclusively operates on empirical evidence, it is a priori excluded that factors that are defined with recourse to co-extensional predicates can be attributed different causal relevancies. If every event instantiates either both of two factors or none of them, the two factors cannot differ with respect to their causal relevance and, therefore, there is no need to distinguish them in causal contexts.

Given the set theoretic analysis of factors in (2.18) and (FI), events can straightforwardly be said to *instantiate* a factor or to be *instances* of a factor iff they are members of the similarity set constituting the corresponding factor. Thus, for all events x :

$$x \text{ instantiates } A \text{ iff } x \in \{x : Ax\}. \quad (2.19)$$

The relation of instantiation holding between an event e and a factor A will, for reasons of brevity, also be symbolically expressed by Ae instead of $e \in \{x : Ax\}$.

Factors shall be negatable. The negation of a factor A will be written thus: \bar{A} . \bar{A} is simply defined as the complementary set of A :

$$\bar{A} =_{df} \{x : \neg Ax\}, \text{ such that } A \text{ satisfies (a), (b), (c), and (d),} \quad (2.20)$$

where "A" is interpretable as any predicate appearing in the definiens of A . Moreover, for all events x :

$$x \in \bar{A} \text{ iff } x \notin A. \quad (2.21)$$

Every event instantiates either A or \bar{A} , not both. A and \bar{A} hence are mutually exclusive sets of events. A shall be labelled a *positive factor*, \bar{A} a *negative factor*. The fact that both a factor A and its negation \bar{A} are non-empty implies that neither $\{x : Ax\}$ nor $\{x : \neg Ax\}$ are all-embracing event sets, i.e. sets that include all events (cf. restriction (b)).

A causal interpretation of negative factors as defined here poses some intricate problems that will be addressed in detail as we proceed. The problems ultimately stem from the fact that negative factors contain *all* events not satisfying a factor defining predicate, which, in ordinary cases, means that the overwhelming majority of events come to be members of virtually any negative factor. However, as we shall see, whenever causal relevance is attributed to a negative factor, only a very specific subset of its instances is of interest at all. For instance, consider anew the car crash example from section 2.2.1. The absence of traffic lights at an intersection can be said to be causally relevant for cars crashing at that intersection. Now take factor

S to be defined as follows:

$$S =_{df} \{x : x \text{ is a signalling of traffic lights}\}$$

We want to say that \bar{S} is causally relevant for car crashes, yet most instances of \bar{S} – as Chancellor Kohl’s promenade on March 2, 1994, or Neil Armstrong’s first human step on the lunar surface – are completely irrelevant for car crashes. What is of interest to an analysis of the causes of car crashes is only the subset of instances of \bar{S} that occur in the spatiotemporal neighborhood of the crashing of two cars. Section 3.6.4 will be concerned with these problems. For now, it suffices to see that caution is called for when it comes to the causal interpretation of negative factors as defined in (2.20).

Factors are of qualitative nature. That is, they are either instantiated by an event e or not instantiated by e . There is no instantiation by degree. Transferred to e.g. the framework of probability theory, factors correspond to binary variables. Nonetheless, factors can involve quantities as the following examples illustrate:

- (49) Storm with winds of over 100 km/h
- (50) Explosion of 10 kg TNT
- (51) Production of 25 mg urea.

Causal enquiries are conducted within a given set of investigated factors. The factors whose causal interdependencies are analyzed may be selected from the set of all factors due to the specific interest of a particular causal investigation or due to epistemological constraints, as the selected factors being known while the unselected factors being unknown. The set of selected factors shall be referred to as the *factor frame* of the corresponding causal enquiry. The members of the factor frame are called *atomic* factors of that frame. The union of atomic factors defines *complex* factors within that frame. Thus, relative to a factor frame consisting of the factors A and B , complex factors C and D are definable as the union or the intersection of A and B :

$$C =_{df} A \cup B \tag{2.22}$$

$$D =_{df} A \cap B \tag{2.23}$$

The term “factor” shall be used as a superordinate term. If not specified, it designates both atomic and complex factors.

Finally, some special relations between factors shall be introduced and defined. Some of these definitions will be equivalently expressed both in terms of first-order predicate logic and of the set theoretic formalism. For reasons of convenience and transparency, we will subsequently mostly make use of the formalism of predicate logic. Two factors A and B are said to be *conditionally dependent* iff (2.24) holds; *conditionally independent* iff the negation of (2.24) holds; *contrary* iff (2.25) holds;

subcontrary iff (2.26) holds; and *contradictory* iff (2.27) holds.

$$\forall x(Ax \rightarrow Bx) \vee \forall x(Bx \rightarrow Ax) \quad \text{resp.} \quad A \subseteq B \vee B \subseteq A \quad (2.24)$$

$$\neg \exists x(Ax \wedge Bx) \quad \text{resp.} \quad A \cap B = \emptyset \quad (2.25)$$

$$\forall x(Ax \vee Bx) \quad (2.26)$$

$$\forall x(Ax \equiv \neg Bx) \quad (2.27)$$

Since complex factors, as atomic factors, are neither empty nor all-embracing event sets, no complex factor is definable by means of contrary, subcontrary, or contradictory factors.

2.3 Causal Relevance

One of the main goals of a theory of causation consists in an analysis of causation – both on type and on token level – in terms of (less controversial or better understood) non-causal notions. It is supposed to provide necessary and/or sufficient conditions for a process to be of causal nature. A theory of causation is expected to clarify the truth-conditions of causal statements as “*a* causes *b*” and “*A* is causally relevant to *B*”. However, our pre-theoretic understanding of causation and, accordingly, our informal assessment of the truth-conditions of causal statements is not consistent. For instance, there are strong pre-theoretic intuitions to the effect that causation is a transitive relation. Nonetheless if a person’s finger is bitten off by a dog and then restored by surgery such that, some years later, the finger regains its full functionality, we would intuitively hesitate to refer to the dog bite as a cause of the perfectly healthy finger, even though the functionality of the finger is certainly the result of a causal chain involving the dog bite.⁸² Or singular causation is often claimed to be asymmetric, nonetheless we are not willing to deny the existence of reciprocal feedback processes on token level – as e.g. in static equilibria – on a priori conceptual grounds. Moreover, on the one hand, we pre-theoretically take causes to – in some sense or other – determine their effects. On the other hand, we are not prepared to abstain from a causal interpretation of processes – as atom bomb detonations – that are irreducibly indeterministic, at least as far as the standard interpretation of quantum mechanics is concerned.

In order to avoid inconsistencies, thus, a theory of causation cannot aim to capture all pre-theoretic intuitions with respect to the characteristics of causation and the truth-conditions of causal statements. Rather, an analysis of causation should only seek to mirror a maximally comprehensive proper subset of pre-theoretic intuitions. In this sense, this whole chapter on the conceptual fundament can be seen as a clarification of the pre-theoretic understanding of causation which shall be theoretically captured in what follows. While the previous section has clarified the analysandum of our upcoming analysis by specifying the relata of causation, this and the following section are going to specify the relational and metaphysical properties of the object of the theory of causation to be developed in chapter 3.

⁸² Cf. Kwart (2001), p. 389.

2.3.1 Singular vs. General Causation

Causal Statements

Singular and general causation are not independent relations. The truth-values of statements expressing causal dependencies on token level can be clarified by means of the truth-values of corresponding statements expressing type level dependencies or vice versa. Consequently, subject to theoretical preferences, analyses of causation commonly explicate one of the two relations in terms of the other. Transference theoretic and counterfactual accounts ordinarily take singular causation to be the basic relation, whereas regularity theories as well as probabilistic or manipulability accounts traditionally consider general causation to be the primary analysandum and accordingly event types to be the primary causal relata.⁸³ The main reason for this preference for event types lies in the non-repeatability of singular events. For regularity theories and for probabilistic accounts repeatability of causes and reproducibility of effects are of crucial importance to an analysis of causation. These theories operate within the Humean tradition that professes the unobservability of causal dependencies in single sequences of events. There is no such thing as an observable or measurable causal tie that bonds causes and effects together. It is only the repeated observation of type-identical event sequences that allows for a causal structuring of the processes under investigation.

The regularity account of causation to be advanced subsequently thus analyzes singular causation with recourse to general causation. In order to illustrate this analytical strategy, a necessary condition for *a* to be a cause of *b* shall be anticipated at this point: Two events *a* and *b* are related as cause and effect only if factor *A*, that is instantiated by *a*, is causally relevant for factor *B*, that is instantiated by *b*. Causal relevance is the central notion to be analyzed in the following. It constitutes the primary object of investigation of the study at hand: *causal structures*. A causal structure can be seen as the set of causal relevance relations holding among the factors in a given frame.

There are some important differences and commonalities among singular and general causation which our upcoming analysis is going to have to theoretically capture. In order to illustrate the salient characteristics of the two causal relations, let us first consider (52) and (53) as paradigmatic statements of singular and general causation, respectively:

(52) *a* caused *b*.

(53) *A* is causally relevant for *B*.

Over and above the obvious differences between (52) and (53) to the effect that (52) is about events, while (53) is about factors, or to the effect that different causal predicates are involved in (52) and (53), there is one important difference between these two statements that deserves separate emphasis. Someone stating

⁸³ For a concise overview over the main theories of causation cf. Baumgartner and Graßhoff (2004).

(52) thereby implies that an event a and an event b have actually taken place such that a was a cause of b . Singular causal statements are implicitly relativized to the spatiotemporal locations of a and b . In contrast, someone stating (53) by no means implies that A or B have been instantiated in a particular situation nor that some specific instance of A caused some specific instance of B . (53) can be true even if instances of A in certain contexts do not cause instances of B . Statements of general causation are not to be understood relative to concrete situations or coordinates in time and space. As mentioned above, the truth of (53) is a necessary condition of the truth of (52). (52), however, might well be false without the truth-value of (53) thereby being affected. Postponing the analysis of causal relevance, only a negative condition as regards the truth of (53) can be offered here: (53) is false if there are *no* instances of A that cause instances of B . For now and with recourse to the notoriously vague notion of a *ceteris paribus* condition⁸⁴, (53) can be informally understood as saying that instances of A *ceteris paribus* cause instances of B .

Moreover, (52) and (53) have one thing in common that needs to be accentuated here. Neither (52) nor (53) identify so-called *full causes*.⁸⁵ That means, (52) does not state that a was the only cause of b , and (53) does not assess that whenever A is instantiated, an instance of B occurs. (52) is to be read as saying that a was *one of* the causes of b , and (53) correspondingly declares that A is *one of* the causally relevant factors of B . Causes are always part of complex structures. A single cause only brings about its effect in combination with other causes. Whenever a full cause is referred to in the following, this shall be explicitly indicated by labels as a *causally sufficient cause* or a *complete/full cause*.

Relational Properties

Apart from these differences and commonalities among statements expressing singular and general causal dependencies, there are some important differences and commonalities between the two causal relations that deserve separate mention. Some of the relational properties of singular and general causation are uncontroversial others are not generally agreed upon. As indicated in the introductory remarks to this section, I take the controversies as regards the relational properties of singular and general causation to stem from inconsistencies in our pre-theoretic causal intuitions. Either causation is asymmetric or it is not, either it is transitive or it is not. However, there are pre-theoretic intuitions that profess asymmetric and transitive causal dependencies, while others see causation to be neither characterized by asymmetry nor by transitivity.

Let us begin with uncontroversial relational properties of singular and general causation, respectively. While general causation is *non-reflexive*, singular causation is *irreflexive*. A relation C is said to be non-reflexive iff (2.28) holds and

⁸⁴ Cf. section 1.2.2 above. Section 5.2 will introduce the notion of *causal homogeneity* as a modern substitute for vague *ceteris paribus* provisos.

⁸⁵ Cf. Mackie (1974), p. 64.

irreflexive iff (2.29) holds.⁸⁶

$$\neg\forall x Cxx \quad (2.28)$$

$$\forall x \neg Cxx \quad (2.29)$$

On type level there exist many causal feedbacks or cycles such that a factor is (directly or indirectly) causally relevant to itself. For instance, rain is causally relevant to flooding which, after the water has evaporated, gives rise to the formation of new rain clouds and, eventually, new rain showers.⁸⁷ Or the influenza infection of a child is causally relevant to the flu of its mother which, again, is causally relevant to a new infection of the child.⁸⁸ While there are causal cycles on type level, the overwhelming majority of type level causal dependencies are not cyclically structured. Thus, many factors are not causally relevant to themselves. On the other hand, no token event ever causes itself. The child's first influenza infection differs from its second infection and concrete rain showers never cause themselves. While cause and effect events may instantiate the same factors, as in cycles, events are always caused by different events. Thus, singular causation is irreflexive as expressed in (2.29). Such as to facilitate later reference to this important feature of singular causation, it shall be labelled:

Irreflexivity of singular causation (NSC): Singular causation, i.e. the relation "... cause/d ...", is irreflexive. No event causes itself.

As indicated in section 1.2.5, both singular and general causation lack symmetry. While for general causation the exact form of this lacking symmetry is generally agreed upon, in case of singular causation matters are controversial. A relation can lack symmetry in several ways, two of which are particularly important for the context at hand: A relation C is non-symmetric iff (2.30) holds and asymmetric iff (2.31) holds.⁸⁹

$$\neg\forall x\forall y(Cxy \rightarrow Cyx) \text{ or equivalently } \exists x\exists y(Cxy \wedge \neg Cyx) \quad (2.30)$$

$$\forall x\forall y(Cxy \rightarrow \neg Cyx) \quad (2.31)$$

In view of the existence of many acyclic causal structures on type level, general causation clearly is non-symmetric. Malfunctioning traffic lights are causally relevant to accidents at intersections, while the accidents are in no way relevant to the defective lights. That means, " A is causally relevant to B " does not imply " B is causally relevant to A ". Moreover, as we have seen above, there are feedback structures on type level. Rainfall may be causally relevant to flooding which is causally relevant to further rainfall. Thus, it may well be the case that A is causally relevant to B and B is causally relevant to A . In consequence, general causation is

⁸⁶ Cf. e.g. Lemmon (1978 (1965)), pp. 183-184).

⁸⁷ Cf. von Wright (1974), pp. 11-12.

⁸⁸ Cf. Davis (1988), p. 146.

⁸⁹ Cf. e.g. Lemmon (1978 (1965)), pp. 180-182.

not asymmetric, i.e. “*A* is causally relevant to *B*” does not imply “*B* is not causally relevant to *A*”.

Singular causation clearly is non-symmetric as well. “*a* causes *b*” does not imply “*b* causes *a*”, as there exists an abundance of causal processes on token level that do not feature feedbacks. Shamus’ striking of a match at a certain time causes the match to light moments later without the lighting of the match in any way causally influencing its being struck by Shamus. However, when it comes to the existence of feedbacks on token level opinions diverge. Candidates for token level feedbacks are e.g. static equilibria. One of the most discussed examples of this kind is the following: Consider a stable house of cards. The two cards at the top lean against each other. Each prevents the other from falling over.⁹⁰ In view of the broad event notion developed in section 2.2.2, the leaning of one card against the other may well be considered an event. And it in fact appears that one of these leaning events causes the other and vice versa. The debate as to the causal nature of static equilibria shall not be entered into. Whatever the outcome will be, it seems plain that the question as to whether the card example constitutes a case of token level feedback or not is not to be decided on a priori conceptual grounds. The existence of simultaneous reciprocal singular causation is to be determined synthetically. In the present context, it shall therefore not be presupposed that singular causation is asymmetric, regardless of the fact that the vast majority of token level causes are not reciprocally caused by their effects.

Non-symmetry of general and singular causation (NSY): General causation, i.e. the relation “... is causally relevant to ...”, and singular causation, i.e. the relation “... causes/d ...”, are non-symmetric.

The non-symmetry of both singular and general causation is what is often referred to as the *direction of causation*. In order to theoretically capture the direction of causation – either in its token- or type level variant –, many accounts of causation are complemented by a condition to the effect that causes temporally precede their effects.⁹¹ Such an account, however, not only gives rise to non-symmetric, but moreover to asymmetric causal relations, for temporal priority is asymmetric. Yet, we have seen above, that, firstly, the existence of feedbacks on type level is undisputed and, secondly, token level feedbacks cannot be excluded on conceptual grounds. Thus, accounting for the direction of causation by means of an asymmetry as temporal priority seems inadequate. Furthermore, Reichenbach (1956) has argued in favor of a conceptual priority of the causal relation. He claims that the direction of time should be analyzed in terms of the direction of causation, not vice versa.⁹² As we have seen in the introduction, regularity accounts of causation have

⁹⁰ Cf. e.g. Pollock (1976), p. 173, Fair (1979), p. 230, Frankel (1986), and Hausman (1998), pp. 44-47.

⁹¹ A very explicit example of this analytical strategy is Suppes (1970).

⁹² Arguments to the same effect can be found in Mellor (1981), ch. 9, or Papineau (1985), pp. 273-274, see also Baumgartner and Graßhoff (2004), pp. 110-113.

often been accused of being incapable of capturing the direction of causation in atemporal terms. It shall be shown in chapter 3 that this criticism is rash.

Finally, there is a lot of discussion in the literature as to the transitivity of causation. A relation C is transitive iff (2.32) holds and non-transitive iff (2.33) holds.⁹³

$$\forall x\forall y\forall z(Cxy \wedge Cyz \rightarrow Cxz) \quad (2.32)$$

$$\exists x\exists y\exists z((Cxy \wedge Cyz) \wedge \neg Cxz) \quad (2.33)$$

Lewis (1973), Lewis (2000), or Hall (2000), for instance, claim causation to be transitive, authors as Eells (1991), Ehring (1997), Kvart (2001), and Hitchcock (2001) have argued in favor of its non-transitivity. Unfortunately, the transitivity debate is diluted by a lot of confusion as to whether the object of controversy is singular or general causation. It does not always become sufficiently transparent whether token or type level causation is claimed to be transitive or non-transitive. It seems to me, however, that most authors denying the transitivity of causation argue against the transitivity of singular causation, the transitivity of general causation largely being accepted.

Intuitively it seems indisputable that, given that an event a causes an b and b causes and event c , a causes c as well. In the course of most of our everyday exposure to causal processes we implicitly assume the transitivity of token level causal dependencies. A farmer only fertilizes his fields because he takes it for granted that this action will cause his crop to be enhanced, which, in turn, will cause his income to be increased. Had the farmer reason to expect that somewhere along this chain the causal impact of the fertilization could spontaneously vanish, he would abstain from the arduous field work in the first place. Causes and effects are connected in chains such that the first event on the chain (indirectly) causes the last. All the authors challenging this intuition argue from concrete examples of token level chains, for which it does not seem to be the case that the first event causes the last. Take, for instance, the fireplace example discussed in Woodward (1984) or Ehring (1997), p. 76: a shall be the event of Shamus putting potassium salts in a fireplace, b the event of a purple fire in the fireplace a little later, and c the event of the neighboring woods being on fire. The resulting forest fire appears to be the result of a causal chain from a to b to c . Yet, even though a causes b and b causes c we would not want to say that a causes c . To refute this example it has been argued that the process leading from a to c does not really constitute a causal chain.⁹⁴ a does not cause that particular property of event b that later causes the wood fire. a causes the color of the fire, yet it is its high thermal radiation that subsequently causes the wood fire. Thus, the causal process at hand should be modelled in terms of a causing b_1 and a different event b_2 causing c .

⁹³ Cf. e.g. Lemmon (1978 (1965)), pp. 182-183. Non-transitivity must be distinguished from *intransitivity*. A relation C is intransitive iff $\forall x\forall y\forall z(Cxy \wedge Cyz \rightarrow \neg Cxz)$. Nobody claims that causation is intransitive.

⁹⁴ Cf. Baumgartner and Graßhoff (2004), pp. 53-59.

While this example might be rejected for not constituting a causal chain, the following example is much more intricate: Take *a* to be the event of a dog biting off Shamus' left pinky, *b* shall be the event of the pinky being restored by surgery, and *c* the pinky's full functionality some years later.⁹⁵ There is no doubt that *c* is the result of a causal chain that is set off by *a*. *a* is a cause (among others, of course) of *b* and *b* is a cause of *c*. Nonetheless *a* does not seem to be a cause of *c*. There are many analogous examples of this sort.⁹⁶ For instance, the 'boulder' case as described in Hitchcock (2001), p. 276:

(...) a boulder is dislodged, and begins rolling ominously toward Hiker. Before it reaches him, Hiker sees the boulder and ducks. The boulder sails harmlessly over his head with nary a centimeter to spare. Hiker survives his ordeal.

The boulder being dislodged does not seem to cause Hiker's survival. The structure of these examples is always the same: a causal chain whose first element does not appear to cause its last. Still, intuitions waver. In case of the dog bite example it might be argued that the dog bite does not seem to cause the fully functional finger because the overwhelming majority of healthy fingers has a dog bite free causal past. While, normally, healthy fingers are not caused by dog bites, in extraordinary cases it might nonetheless happen that dog bites causally bring about the full functionality of fingers. Thus, what at first sight appears to be an example demonstrating the non-transitivity of singular causation, in fact is just a rare causal chain whose first element causes its last as in every other causal chain. Similar explanations of the seemingly lacking causal influence of the dislodged boulder on Hiker's survival might be found. Accordingly, these examples are contentious.

Yet, the two relational properties we have thus far attributed to singular causation – irreflexivity and non-symmetry – constitute substantial further grounds against the transitivity of that relation. There is no self-causation on token level and in light of static equilibria we did not exclude reciprocal causal feedbacks for mere conceptual reasons. Given these two properties, singular causation cannot be transitive, for every irreflexive and transitive relation is asymmetric.⁹⁷ That means, taking singular causation to be transitive would require either allowing for self-causation or stipulating that there are no causal feedbacks on token level. None of these maneuvers to secure the transitivity of singular causation are outright impossible. Determining the relational properties of causation is a matter of balancing incompatible, yet underdetermined intuitions. There are no decisive arguments for or against ascribing one property rather than another. I shall ascribe relational properties to causation according to the following principle: Whenever there is reasonable doubt that singular or general causation feature a particular relational

⁹⁵ Cf. Kwart (2001), p. 389.

⁹⁶ An excellent overview over the hard cases against transitivity of singular causation can be found in Hall (2000).

⁹⁷ Cf. Lemmon (1978 (1965)), p. 185.

property, abstain from ascribing it to the corresponding causal relation. Therefore, as in case of the doubtful asymmetry of singular causation, I shall in case of transitivity again abstain from attributing a doubtful property to the token level causal relation on mere conceptual grounds. I hence assume singular causation to be non-transitive.

In case of general causation, however, matters are different. Factors may well be causally relevant to themselves. If A is causally relevant to B and B is causally relevant to A , it makes perfect sense to hold that A is causally relevant to A . Rainfall is causally relevant to flooding which is causally relevant to further rainfall, thus rainfall is causally relevant to rainfall. While there is no self-causation on token level, self-relevance is ubiquitous in causal cycles. The only example that seems clearly designed to demonstrate the non-transitivity of general causation is the famous birth-control example that can e.g. be found in Hesslow (1976). Taking contraceptive pills, factor A , is causally relevant to the presence of a thrombosis-causing chemical in the blood, B , which is causally relevant to thrombosis, C . At the same time, however, A is causally relevant for not being pregnant, \bar{D} , which in turn is causally relevant to not suffering from thrombosis, \bar{C} . The idea behind this example is that, while A is causally relevant to B and B is causally relevant to C , the effect of A on C is cancelled out by the causal path involving \bar{D} .

At most, the birth control case demonstrates that a probabilistic theory of causation does not yield a transitive notion of causal relevance, for even though A raises the probability of B and B raises the probability of C , A does not raise the probability of C . The core analysis of a probabilistic theory of causation – “...raises the probability of...” – is not transitive. However, a deterministically conceived notion of general causation – as the one to be analyzed in this study – is not shown to be non-transitive by the birth-control example.⁹⁸ If A determines B to be instantiated and B determines C to be instantiated, A , of course, determines C to be instantiated. Unless a factor is contradictory, it does not determine C and \bar{C} . If in a concrete case someone takes the pill and does not get thrombosis, say because the development of thrombosis is simultaneously prevented on a different causal path, factor A is not deprived of its causal relevance to C . Such a case, rather than proving the irrelevance of A , simply shows that A is not sufficient for C in isolation, but part of a complex cause of C . In order for A to be causally relevant to C it takes at least one scenario in which an instance of A causes an instance of C . Thus, positive and negative causal relevance do not cancel out and, accordingly, causal relevance, i.e. general causation, can be presumed to be transitive.

Transitivity of general causation (TR): If a factor A is causally relevant to a factor B and B is causally relevant to C , A is (indirectly) causally relevant to C .

It needs to be stressed again that pinning down the relational properties of singular and general causation in this stipulative manner involves a certain arbitrariness. It is a matter of assessing inconsistent pre-theoretic intuitions, of preferring

⁹⁸ Cf. Hesslow (1976), Eells (1991), pp. 211-222.

some intuitions to others. This could be done in several ways without there being a decisive argument for or against any of these pre-theoretic notions of causation. Nonetheless, it is inevitable that the relational properties ascribed to singular and general causation are clarified in some way or another, for, as mentioned above, pre-theoretic intuitions with respect to causation are not consistent. It cannot be the goal of a theoretical account of causation to mirror all informal assessments of the truth-conditions of causal statements. The driving force behind the above ascription of relational properties to singular and general causation has been theoretical caution: Whenever there was a reasonable doubt that one of the causal relations features a certain relational property or not, I abstained from ascribing it to the corresponding relation.

After having clarified the most notable relational properties of general causation by contrasting it with singular causation, we are now going to further pave the ground for our analysis of causal relevance by reviewing the elementary causal structures.

2.3.2 Graphical Notation

Causal relevance is illustratively represented by graphical means. Graphical representation of causal relevance structures has been given a detailed discussion in Baumgartner and Graßhoff (2004).⁹⁹ At this point, I shall hence only provide a short overview over the central conceptual distinctions in regard to causal graphs – with central definitions contrasted against the continuous text.

Causal graphs are commonly taken to be directed graphs, or *digraphs* for short.¹⁰⁰ Digraphs consist of two basic elements: vertices and directed edges. The vertices of causal graphs represent factors, the edges stand for ordered pairs of factors. (53) is expressed by the digraph (a) in figure 2.1. A and B are said to be *neighbors* such that A *dominates* B . The edge $\langle A, B \rangle$ *leaves* A and *enters* B . A is its *tail* and B is its *head*. The edge $\langle A, B \rangle$ is also called a *path*.

Path: A sequence of vertices $V = Z_1, Z_2, \dots, Z_k, k \geq 2$, constitutes a path from Z_1 to Z_k iff for each Z_i and $Z_{i+1}, 1 \leq i < k$, in V : The pair $\langle Z_i, Z_{i+1} \rangle$ is a directed edge oriented from Z_i to Z_{i+1} .

Furthermore, relative to the set of factors contained in graph (a) of figure 2.1, A is called a *root factor*.¹⁰¹

Root factor: Relative to a causal graph G , a factor $A \in G$ is a root factor of G iff A appears in G only as tail, not as head.

If not explicitly indicated to the contrary, causal graphs are *incomplete* in the sense that they do not represent causally sufficient causes or full causes. They select

⁹⁹ Cf. Baumgartner and Graßhoff (2004), pp. 59-67.

¹⁰⁰ For graph theory in general see e.g. Bang-Jensen and Gutin (2001) and for causal digraphs in particular Spirtes, Glymour, and Scheines (2000 (1993)).

¹⁰¹ Probabilistic accounts of causation would speak of A in terms of an *exogenous variable*.

a subset of all causally relevant factors of a given effect, most likely the subset of factors that are known or that are of interest to a particular causal investigation.

Finally, in order to minimize the amount of edges in a causal graph, its edges are taken to represent the relation of *direct* causal relevance, which is a special case of causal relevance and which will be introduced in the next section.

2.3.3 Direct vs. Indirect Relevance

As events can be causally ordered in chains such that certain events are direct and others indirect causes, a factor A can be *directly* or *indirectly* causally relevant for a factor B . In accordance with the terminological conventions devised on page 57, the notions of a *direct* and of an *indirect cause* will be applied both to causing events and to causally relevant factors. In graph (a) of figure 2.1 A is a direct cause of B . As factor B in graph (b) illustrates, vertices in causal graphs can be both tails of one edge and heads of another. Ordered pairs of causally related factors can combine to form chains. In (b) A is not a direct, but an indirect cause of C . Whether the causal relevance of A for C is direct or indirect depends on there being a third event B within the respective causal structure such that A is causally relevant for B and B is causally relevant for C .

Direct cause: Relative to a causal graph G , a factor $A \in G$ is a direct cause of a factor $B \in G$ iff there is an edge $\langle A, B \rangle$ in G .

Indirect cause: Relative to a causal graph G , a factor $A \in G$ is an indirect cause of a factor $B \in G$ iff G contains factors $Z_1, Z_2, \dots, Z_k, 1 \leq k$, such that the sequence A, Z_1, \dots, Z_k, B constitutes a path in G .

Direct and indirect causal relevance will be given precise analyses later on. Three things need to be called attention to at this point already. First, the causal relevance of a factor is direct or indirect only relative to the level of specification chosen and the totality of factors analyzed in the course of a causal investigation. A may be directly causally relevant relative to one set of factors and indirectly relevant relative to another set. Second, a factor A may be both directly and indirectly causally relevant to B . Consider the following scenario: By mistake Tom eats a poisonous mushroom. The resulting gastrospasm causes him to have cramps in his back. The next morning Tom does not go to work due to the gastrospasm and due to the cramps in his back. In this scenario Tom's gastrospasm instantiates a factor (gastrospasm) which is both directly and indirectly causally relevant to a temporary inability to work. It is directly causally relevant inasmuch as its relevance is not mediated by an intermediary factor on one path, while it is indirectly causally relevant inasmuch as its relevance is mediated by a backache on

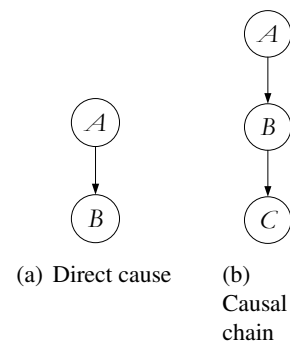


Fig. 2.1: Direct and indirect causal relevance.

another path. Third, the notion of causal relevance is superordinate to direct and indirect relevance. That means, a factor A is causally relevant to a factor B iff A is directly or indirectly relevant to B .

2.3.4 Complex Causes and Coincidences

Events virtually never bring about an effect in isolation, i.e. independently of other events. Correspondingly for factors: There are no factors that are alone in being causally relevant for a given effect. For a match to light it takes much more than it merely being struck. Oxygen needs to be present, the match needs to be dry etc. Single events and factors, in order to be causally effective, are parts of whole causing complexes, or *complex causes*. Such as to graphically represent complex causes, the standard graphical symbolism needs to be extended. We represent a causal structure where instances of A and B bring about an instance of C only if both A and B are instantiated in a given situation – as illustrated in figure 2.2 – by an arch connecting the edges $\langle A, C \rangle$ and $\langle B, C \rangle$ as they enter their common head.

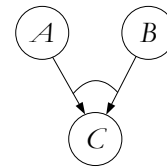


Fig. 2.2: Complex cause.

Every factor of a complex cause is an indispensable part of that cause, i.e. whenever one of these factors is not instantiated, the corresponding effect does not occur due to the respective complex cause. This implies that factors of a complex cause are not conditionally dependent. No factor of a complex cause is a subset of another factor in that complex cause.¹⁰² If the instantiation of A in figure 2.2 determines another factor A' to be instantiated as well, A' cannot be part of the complex cause depicted in figure 2.2. Take A to be the striking of a match and A' to be the striking of a match against a matchbox. These two factors are conditionally dependent: $A' \subseteq A$. Whenever A' is instantiated, A is instantiated as well. It is however possible that A is instantiated without an instance of A' such that C is instantiated nonetheless, for instance, when the match is struck against a wall. Hence, given A , A' is dispensable for bringing about C and therefore not part of the complex cause that includes A . Still, A' is part of another complex cause of C – of a complex cause that besides A' includes factors as presence of oxygen and dryness of the striking surface etc., all of which are conditionally independent of A' . Moreover, in the next section we shall see that factors, in order to be attributable causal relevance, must prove to be indispensable for the instantiation of the corresponding effect in at least one test situation. That means, causes are instantiatable. Therefore, factors constituting a complex cause may neither be contradictory nor contrary. A and \bar{A} are not co-instantiatable, for no factor can be present and absent at the same time.

¹⁰² Conditional dependence between factors is defined in (2.24) on p. 65.

A complex cause only becomes causally effective if all of its constituents are co-instantiated. This raises the important question as to how the notion of co-instantiation is to be understood in this context. Consider anew the example in figure 2.2. *A* shall again be the striking of a match and *B* shall be interpreted as the presence of oxygen. If a match is struck in a vacuum chamber and oxygen is present outside the chamber, both *A* and *B* are instantiated. Can *A* and *B* be said to be co-instantiated in that situation? They definitely are not co-instantiated such that the match actually catches fire. However, they are co-instantiated in the sense that both *A* and *B* are instantiated in this particular situation. Apparently thus, for factors of a complex cause to be instantiated in a way for those instances to be causally effective it takes more than a mere conjunction as “*A* and *B* are instantiated” being true. Those instances must be related in a very specific way. While this relationship clearly consists in some form of spatiotemporal proximity, the exact spatiotemporal interval required for proper co-instantiation of members of a complex cause is far from evident. In the above scenario, for instance, *A* and *B* are temporally coincident and can easily be assumed to be only fragments of a millimeter apart – separated by the ultra thin glass of the vacuum chamber. The reason for the match not catching fire, evidently, is the struck match not being directly surrounded by oxygen. This seems to suggest that spatiotemporal proximity, in order for instances of a complex cause to become causally effective, must be understood in terms of spatiotemporal contiguity: Only temporally coincident and spatially directly conjoined instances of a complex cause seem to be properly related to each other such as to become causally effective. That, however, would be asking too much. For consider the following context: Assume that drunken driving, factor *D*, and slipperiness of the street, *E*, together are causally sufficient for a car accident. Now, take a concrete situation in which drunken Tom drives through the rain and due to this co-instantiation of *D* and *E* Tom crashes into a nearby tree. In the course of this incident *D* and *E* are co-instantiated without actually being in direct spatial contact – the drunkenness being located somewhere in Tom’s body and the slipperiness of the street being located below Tom’s car. The spatiotemporal relation subsisting between causally effective instances of a complex cause varies from process to process and depends on the level of specification chosen to describe causal processes. Sometimes these instances are contiguous, sometimes they are not.

Characterizing the relationship that has to subsist among instances of a complex cause for them to become causally effective will prove to be a very intricate problem. It is a problem that is mostly sidestepped in theories of causation, regardless of the theoretical background chosen. The problem basically consists in an analysis of this relationship in non-causal terms. Ordinarily, the causal relation does not subsist between single instances of causes and effects, but between *co-instances* of a complex cause and the corresponding effect. The definiendum of an analysis of singular causation will thus be of somewhat the following form: Co-instances of a complex cause consisting of the factors Z_1, Z_2, \dots, Z_n cause

an instance of a factor Z_o iff Hence, in order to avoid circularity, the relationship among instances of a complex cause, whose instances are causally effective, cannot be analyzed by presupposing an analysis of the causal relation, i.e. this relationship cannot be defined by declaring it to be such that thus related instances actually become causally effective.

For lack of the theoretical devices that will be developed in the following, the solution of this problem has to be postponed at this point. For the purpose of easy reference to the relationship between causally effective instances of a complex cause, we shall resort to the notion of a *coincidence*. Whenever the factors of a complex cause are instantiated such as to become causally effective, these instances shall be labelled *coincident*. For now, the notion of a coincidence will be left deliberately vague. By a coincidence we shall understand the co-instantiation of at least two conditionally independent factors in suitable spatiotemporal proximity. The fuzzy expression here, of course, is “suitable”. Instances of two factors Z_1 and Z_2 are instantiated suitably proximately iff (1) whenever Z_1 and Z_2 constitute a causally sufficient complex cause of a factor Z_o , Z_o is instantiated due to the co-instantiation of Z_1 and Z_2 , and (2) whenever Z_1 and Z_2 do not constitute a causally sufficient complex cause of a factor Z_o , Z_o is not instantiated due to the co-instantiation Z_1 and Z_2 . A coincidence of factors Z_1 and Z_2 is symbolically represented by a mere juxtaposition of its factors. This yields Z_1Z_2 for the coincidence at hand.

Coincidence: A coincidence $Z_1Z_2 \dots Z_n$ is the co-instantiation of conditionally independent factors $Z_1, Z_2, \dots, Z_n, n \geq 2$, such that

- (1) if Z_1, Z_2, \dots, Z_n constitute a causally sufficient complex cause of a factor Z_o , Z_o is instantiated due to the co-instantiation of Z_1, Z_2, \dots, Z_n .
- (2) if Z_1, Z_2, \dots, Z_n do not constitute a causally sufficient complex cause of a factor Z_o , Z_o is not instantiated due to the co-instantiation of Z_1, Z_2, \dots, Z_n .

That means, a coincidence is the co-instantiation of factors such that these factors become causally effective iff they are thus effective. Or put differently, if instances of factors are coincident, the effects of these factors do not fail to occur due to spatiotemporal incongruence. Clearly, this is not – as required above – a non-causal characterization of the notion of a coincidence. More will have to be said about coincidences later on.¹⁰³ In the meantime, this rough and inappropriate characterization will have to suffice.

2.3.5 Alternative Causes

Effects are not only brought about along one single causal path. Normally there are various different paths that lead to an effect. A match can be lit by striking it

¹⁰³ Cf. p. 97 below.

or exposing it to fire or by means of inflammable chemicals. All these alternative causal paths are independent of each other. Striking a match and exposing it to fire are said to be *alternative causes* of a match catching fire. As illustrated in figure 2.3, alternative causes are graphically represented by edges with different tails and identical heads. In contrast to complex causes, in case of alternative causes there is no arch connecting the edges at their heads.

Corresponding to what has been said about the truth-conditions of causal relevance statements in section 2.3.1, the graph in figure 2.3 states that there is at least one situation in which an instance of A causes an instance of C and one situation in which an instance of B causes an instance of C . These two test situations differ. For in order for A and B to be alternative causes of C , instances of A have to cause events

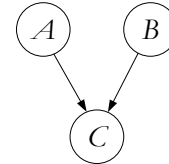


Fig. 2.3: Alternative causes.

of type C independently of instances of B , and vice versa. This independence is only guaranteed if there is at least one situation in which A and \bar{B} are instantiated such that an event of type C is brought about and at least one situation in which \bar{A} and B are instantiated such that an instance of C is realized. For this condition to be satisfied, A and B cannot be conditionally dependent. If A were to include B or B were to include A , A and B could not be alternative causes of C .

Moreover, in order for A and B to be alternative causes of C , A and B are each indispensable for the bringing about of C in at least one test situation. Hence, for A (and B) at least one situation is required, in which A (B) is instantiated along with C , and at least one situation, in which \bar{A} (\bar{B}) is instantiated along with \bar{C} . Due to this second requirement, A and B cannot be subcontrary. Alternative causes are not subcontrary. For were they subcontrary, one of the two alternative causes would always be instantiated and, hence, there would not be any test situations, in which the effect does not occur.

2.3.6 Common Causes

Causes often have multiple effects. An approaching cyclone is both causally relevant for rainfalls and for sinking barometer standings. Short-circuits are causally relevant for defect TV sets, defect light bulbs, and for fires breaking out. Causes that have multiple effect are called *common causes* of these effects. Common causes are graphically represented by edges with common tails and different heads. Common cause structures, as depicted in figure 2.4, are also referred to as *epiphenomena*.¹⁰⁴ In figure 2.4 A is the common cause of B and C .

¹⁰⁴ Note that this terminology differs from the notion of an epiphenomenon used in the literature on mental causation. In the latter context an epiphenomenon is a physically caused mental side effect which itself cannot cause anything. Here “epiphenomenon” just describes a causal structure as depicted in figure 2.4. No assumptions as to the causal impotence of either B or C are implied by referring to such a structure as being *epiphenomenal*.

It needs to be stressed that a cause having multiple effects is not sufficient for that cause to be a common cause of these effects. Factor A in graph (b) of figure 2.1 has multiple effects, without being a common cause of these effects. A cause is a common cause of its multiple effects only if neither of these effects is causally relevant to the other effect. That means, among the multiple effects of an epiphenomenon there is no causal dependence.

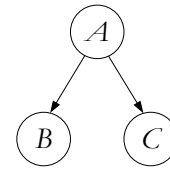


Fig. 2.4: Common cause.

2.3.7 Cycles

As mentioned before, causal dependencies on type level can be cyclically structured. Many economical or biochemical processes have a form as depicted in figure 2.5. While causal relevance relationships can be cyclically structured, events

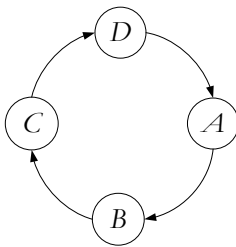


Fig. 2.5: Causal cycle.

never cause themselves. There is no self-causation on token level. Every time a factor A is instantiated in the course of a process featuring a causal structure as in figure 2.5 it is instantiated by *different* events a_1 , a_2 , etc. Furthermore, causal cycles constitute the only causal structure without root factors. In a cycle every factor dominates every other factor, including itself. This multitude of dependencies among the factors in a cycle substantially limits the diversity of combinations in which corresponding factors can be co-instantiated.

The four factors involved in the cycle of figure 2.5, for instance, may only be jointly present or absent. That means, $ABCD$ and \overline{ABCD} are the only two empirically possible coincidences, given that the behavior of these factors is in fact regulated by the cycle in 2.5. All acyclic causal structures generate much more diverse empirical data, which greatly facilitates their experimental or observational detection.¹⁰⁵ The severely limited diversity of empirical data generated by causal cycles has led to their almost general disregard within theoretical accounts of causation. All currently discussed theories of causation and methodologies of causal reasoning explicitly or implicitly restrict their validity and applicability to acyclic causal structures, the most prominent example clearly being Spirtes, Glymour, and Scheines (2000 (1993)). Even though I hold that the regularity account proposed subsequently could be adapted for a successful treatment of causal cycles, I shall follow the usual practice in the present study and bypass the problems imposed by causal cycles. To properly account for causal cycles within a regularity theoretic (or any other) framework requires substantial additional work. I shall thus focus on the correct analysis of complex causal structures composed of the other atomic structures discussed in this section only.

¹⁰⁵ The enhanced detectability of structures that generate diverse data is illustrated and substantiated in chapter 5.

2.4 Causal Principles

While the previous section has clarified the relational properties of the object of the theory of causation to be developed in chapter 3, the section at hand further specifies our analysandum by introducing four general principles that are claimed to be satisfied by the causal relation we are going to analyze in the next chapter. The first two of these principles define our analysandum as being a deterministic relation, while the third and fourth principle stipulate that the causal nature of a process must be determinable by empiricist Humean means. Especially the deterministic character of causation, of course, is a matter of vehement metaphysical controversy, which, however, will not be picked up in the present context. The study at hand assumes causation to be deterministic. Or, less radically, if there in fact are irreducibly indeterministic causal contexts, the present analysis simply does not apply to them.¹⁰⁶

2.4.1 Principle of Determinism

The Principle of Determinism is the most widely known among the four principles. It captures the idea that causes fix their effects. More precisely, it states that in a situation S_1 , that accords with another situation S_2 as regards the instantiation of root factors, type-identical effects are instantiated as in S_2 .

Principle of Determinism (PD): In any two situations that accord with respect to instantiations of root factors type-identical effects are instantiated. *Same causes, same effects.*

This principle expresses the core of a deterministically conceived causal relation. There are no such things as alternative effects. Instances of root factors do not cause one set of effects on one occasion and another set on a different occasion. The Principle of Determinism can be given both a strong and weak reading with respect to the consequences drawn from it. If, as claimed by the standard interpretation of quantum mechanics, there are irreducibly probabilistic processes, the strong reading interdicts their causal interpretation. The weak reading, on the other hand, could allow for a causal interpretation of processes, yet determines that the analysis of causation proposed in the present context is inapplicable to these scenarios. The weak reading thus simply constrains the domain of the theory of causation at hand to macro level events and assumes possible indeterminacies on micro level not to ‘percolate up’ to the macro level. It should be stressed again that even the most influential theories of probabilistic causation confine the applicability of their accounts to so-called pseudo-indeterministic structures, which are nothing but

¹⁰⁶ Cf. section 1.2.8.

incompletely specified or known deterministic structures.¹⁰⁷ Probabilities, according to these accounts, mirror epistemic uncertainty rather than hard-boiled physical indeterminacies.

According to one of the most often advanced objections against regularity accounts of causation, presupposing (PD) implies the world to be deterministically structured on mere conceptual, i.e. a priori, grounds. Yet, the question as to the deterministic structuring of the world, thus the objection continues, is not to be decided a priori, but is a matter of scientific investigation.

If we do not know *a priori* whether the world is deterministic, our analyses of causality have no business deciding the matter without reference to the results of science. All of the Humean analyses of causality, including Mackie's, have simply *assumed* that determinism is true, and therefore they are all of them philosophically inadequate.¹⁰⁸

This objection, however, can neither be raised against (PD) in its strong nor in its weak reading. (PD) does not profess the world to be deterministically structured. The existence of indeterministic occurrences is compatible with this principle. The principle, if given the strong reading, just precludes these indeterministic occurrences from a causal interpretation, if given the weak reading, it excludes these occurrences from a treatment by the theory of causation to be developed here. Every causal process in the domain of our theory satisfies the Principle of Determinism. In this sense, rather than ruling the world to be deterministic on a priori grounds, (PD) is merely a principle that constitutes part of the explication of the *notion* of causality presupposed in the present context.

2.4.2 Principle of Causality

The second principle devises effects not to occur spontaneously. All events involved in a causal process are causally generated.

Principle of Causality (PC): If no cause is instantiated, no effect is instantiated.

This is not to be confounded with the *Law of Universal Causation*¹⁰⁹ according to which every event has a cause. Rather, (PC) states that factors that are interpreted to be effects of other factors are not instantiated without instances of their causing factors. The Principle of Causality also holds for irreducibly probabilistic structures, within which effects may not be determined, nonetheless, they do not occur spontaneously.

(PC) can be seen as an analytical statement that explicates the meaning of the notion of an effect. No event can be claimed to be an effect without a corresponding cause having occurred.

¹⁰⁷ Cf. Spirtes, Glymour, and Scheines (2000 (1993)), pp. 52, 63, Cartwright (1999), Twardy and Korb (2004).

¹⁰⁸ Korb (1999), p. 267.

¹⁰⁹ Cf. Mill (1879 (1843)), book III, ch. 5.

2.4.3 Principle of Relevance

The Principle of Relevance mirrors two of the central features of complex and alternative causes, respectively, that we uncovered in sections 2.3.4 and 2.3.5. Complex causes do not contain redundant elements and each alternative cause has to be indispensable for the bringing about of its effect in at least one test situation. Elements of complex causes must be conditionally independent, they must be co-instantiatable in all logically possible combinations. Similarly for alternative causes: An effect does not have alternative causes that are redundant. In order to illustrate this latter constraint, consider the following example: Suppose two factors A and B are often co-instantiated along with a third factor C and, moreover, we know of A that it is causally relevant to C . Now, a particular situation S_1 is observed in which events a and b , that are instances of A and B , respectively, are followed by event c which instantiates C . Furthermore, it shall be assumed that we are familiar with a set of other test situations S_2 to S_n in which events of type A are co-instantiated with instances of C without concurrent instantiations of B . However, there shall be no test situations to the effect that instances of B are co-instantiated with events of type C while A is absent. Given this constellation, there are no grounds on which to interpret B as an alternative cause of C and, accordingly, to take c in S_1 to be overdetermined by a and b . Causes must demonstrate their relevance in at least one test situation in which all alternative causes of the corresponding effect are absent. While A and B , thus, cannot be alternative causes of C , they might still both be causally relevant to C . B could, for instance, be a cause of A which, in turn, is a cause of C .

In order to account for this kind of parsimony as regards complex and alternative causes, we introduce the Principle of Relevance:

Principle of Relevance (PR): Every causally relevant factor is a non-redundant part of a (possibly complex) cause for which there is at least one test situation in which that cause brings about an instance of the corresponding effect while all alternative causes of that effect are absent.

Or put differently and more specifically: A factor Z_1 is causally relevant to a factor Z_2 only if there is at least one causal path from Z_1 to Z_2 on which Z_1 is indispensable and for which there is at least one test situation in which an instance of Z_1 causes an instance of Z_2 while all alternative causal paths to Z_2 that do not contain Z_1 are not instantiated.

Consider anew the causal graph depicted in figure 2.3. This structure is only consistent with the Principle of Relevance if there is a test situation S_1 , in which A and \bar{B} are instantiated such that the instance of A causes an instance of C , and a test situation S_2 , in which \bar{A} and B are instantiated such that an event of type B causes an instance of C . For obvious logical reasons, S_1 and S_2 cannot be identical.

There are some epistemological restrictions as regards the evaluation of whether a causal structure actually complies with (PR). The principle does not

require the called for test situations to exist in the past or the present of a particular causal investigation. The test situations that establish the indispensability of a causal factor simply need to exist in a timeless sense, that is, in past, present or future. For methodological reasons, however, it is commendable not to integrate a factor into a causal structure unless its indispensability has been proven. Of course, as always with mere methodological guidelines, causal diagnoses based on these guidelines can easily be demonstrated to be incomplete by later discoveries. That a factor *A* does not conform to (PR) at a certain stage of a causal investigation, does not imply that *A* is *causally irrelevant* to the causal structure under investigation. As shown in Baumgartner and Graßhoff (2004), contrary to causal relevance, causal irrelevancies are not derivable from experimental findings.¹¹⁰ It can always happen that test situations turn up that require the integration of a long neglected factor into a causal structure.

2.4.4 Principle of Persistent Relevance

Causal structures, as causal explanations, are typically incomplete. In order to retain sufficiency, a regularity statement can be prefixed by a *ceteris paribus* clause.¹¹¹ Successful and reliable causal diagnoses within such a mainly unknown background require a fourth principle. For an illustrative introduction of this principle, consider the following context: Suppose, factors *A*, *B*, and *C* to constitute a complex cause of an effect *E*. Thus, instances of *A* only bring about events of type *E* if *B* and *C* are coincidentally instantiated as well. Moreover, coincident instances of *A*, *B*, and *C* shall be causally sufficient for *E*. Yet, *A* not being causally sufficient for *E* on its own, shall be unknown to the scientific discipline that investigates this causal structure, i.e. the corresponding discipline is unaware of the causal relevance of *B* and *C*. Regardless of this limited knowledge the causal relevance of *A* is diagnosable in situations in which *B* and *C* are accidentally and unknowingly instantiated along with events of type *A*. Soon, however, important scientific progress is made to the effect that the causal relevance of *B* and *C* is discovered. At this point, the Principle of Persistent Relevance is brought to bear. It stipulates the causal relevance of *A* to persist in light of the discovery of the two new causal factors. A factor that is part of a causal structure does not drop out of that structure upon the discovery of new relevancies. If *A* were no longer attributed causal relevance to *E* due to, say, a violation of the Principle of Relevance in light of the discovery of *B* and *C*, *A* would not have been causally relevant in the first place. According to the Principle of Persistent Relevance, a factor *A* is causally relevant to an effect *E* only if *A* retains its relevance upon any extension of the corresponding factor frame. Or, as Graßhoff and May (2001) put it:

¹¹⁰ Cf. Baumgartner and Graßhoff (2004), p. 212. Causal irrelevancies could only be derived from experimental findings, if (unrealistically) strong causal assumptions as regards the background of an investigated experiment were taken for granted (cf. section 5.3.1, p. 201).

¹¹¹ Cf. sections 1.2.2 and 2.3.1.

Increase in causal knowledge is, in this sense, monotonic.¹¹²

Principle of Persistent Relevance (PPR): Causal relevancies are persistent upon extensions of factor frames.

Prima facie, this principle might appear to be an overt triviality. Why, one might ask, should a reportedly causally relevant factor no longer be attributed causal relevance just because further factors are taken into account? Nonetheless, when it comes to identifying false diagnoses, this principle is of crucial importance. In scientific practice it happens every now and then that factors are shown to be dispensable by the mere discovery of new causes. For instance, nowadays medical sciences are convinced that smoking causes lung cancer. Still, it is possible that a genetic predisposition will be discovered in the future that both causes a predilection for smoking tobacco and an increased risk of lung cancer. Smoking, being a mere epiphenomenon of lung cancer, would thus violate the Principle of Relevance and accordingly be rendered dispensable for the explanation of lung cancer. It would drop out of the causal structure that regulates the emergence of lung cancer. In such a case, the Principle of Persistent Relevance denies smoking the status of a causally relevant factor for lung cancer in the first place.

2.5 Features of Causation

Apart from specifying the nature of causation by introducing four principles, this chapter has revealed a number of important features of singular and general causation. As these features and the principles of causation are going to be theoretically accounted for in the next chapter, they shall be recapitulated and listed before we move on. Our theory of causation has to represent all of these features. Their conjunction therefore can be seen as a necessary condition of the success of our analysis of causation. If a definiens of causal relevance and singular causation, respectively, should turn out to violate any conjunct of that condition, that definiens cannot be considered adequate. In order to enable consecutive reference to each particular feature, they shall all be furnished with a label.

Irreflexivity of singular causation (NSC): Singular causation is irreflexive. No event causes itself.

Transitivity of general causation (TR): General causation is transitive.

Non-symmetry of general and singular causation (NSY): General and singular causation are non-symmetric.

Incompleteness (INCOM): Causal structures typically are incomplete.

Spatiotemporal proximity (PROX): Causally related events are spatiotemporally proximate.

¹¹² Graßhoff and May (2001), p. 90.

Coincident instantiation (COIN): Instances of complex causes cause their effect only if coincident.

Principle of Determinism (PD): In any two situations that accord with respect to instantiations of root factors type-identical effects are instantiated.

Principle of Causality (PC): If no cause is instantiated, no effect is instantiated.

Principle of Relevance (PR): Every causally relevant factor is a non-redundant part of a (possibly complex) cause for which there is at least one test situation in which that cause brings about an instance of the corresponding effect while all alternative causes of that effect are absent.

Principle of Persistent Relevance (PPR): Causal relevancies are persistent upon extensions of factor frames.

The Principle of Relevance (PR) implies a further important characteristic of complex and alternative causes:

Conditional independence (CI): Factors within a complex cause as well as alternative causes are conditionally independent.

3. CAUSAL RELEVANCE ANALYZED – MINIMAL THEORIES

3.1 Introduction

Hume and Mill have influentially argued that there is no feature or property of a single event sequence that would mark it to be of causal nature.¹ A causal interpretation of an event sequence, according to these two godfathers of regularity accounts, is rendered possible only if the events of the sequence instantiate factors which satisfy a material conditional as “Whenever A is instantiated, B is instantiated”. For an event a to be identified as a cause of an event b , it is required that a instantiates a factor A such that instances of A are always followed by events of type B , which is instantiated by b . Causes are thus analyzed to be *sufficient conditions* of their effects.

For various reasons, as shown in the introduction, this analysis is inadequate and overly simple. There in fact are no universal regularities of the required kind in nature. Therefore, regularity statements are traditionally prefixed by a notoriously vague *ceteris-paribus* clause. Yet, if thus prefixed, not only genuine causes are given a causal interpretation, the parallel effects within an epiphenomenon have to be seen as causally dependent as well, for they satisfy a material conditional as do genuine causes and effects. Moreover, material conditionals are monotonic, i.e. their antecedents can *salva veritate* be conjunctively supplemented by further factors that might be completely irrelevant to the effect in question. If striking a match is sufficient for the match to catch fire, the combination of striking a match and singing a song is thus sufficient, too. Additionally, material conditionals are true if their antecedents are false. That means, “Whenever Lake Thun is made of gold, people grow wings” is a true material conditional, yet, of course, we would not want to say that the (inexistent) gold in Lake Thun causes humans to grow wings. However, within the Humean framework, that does not include any restrictions as to the predicates admissible for factor definitions, all empty regularities would have to be given a causal interpretation. Also, what Armstrong (1983) calls single-case uniformities, poses serious problems for a Humean account. Consider the conditional “Whenever Nero sets fire on Rome, the Titanic sinks”. This conditional is trivially true, nonetheless we are not prepared to hold Nero responsible for the sinking of the Titanic. Finally, $A \rightarrow B$ is a true material conditional iff $\overline{B} \rightarrow \overline{A}$ is so too. Which of these true conditionals is to be causally interpreted? It is certainly not the case that a factor is causally relevant to another factor iff

¹ Cf. Hume (1978 (1740)), Hume (1999 (1748)), Mill (1879 (1843)).

the negation of the latter is causally relevant to the negation of the former. Smoking is causally relevant to lung cancer, without the absence of lung cancer having causal impact on abstinence from smoking. Hence, intricate questions as to the non-symmetry of causation arise. All in all, A being a sufficient condition of B in the above simplistic sense is neither necessary nor sufficient for A to be a cause of B .

As causes cannot merely be identified with sufficient conditions of their effects, they are not purely *necessary conditions* of the latter either – as was e.g. suggested by Hobbes.² Effects have multiple alternative causes, neither of them being necessary for the occurrence of their effects.³

However, the fact that causes are not identifiable with sufficient or necessary conditions does not vitiate the basic analytical strategy of a regularity theory, that, as its name suggests, consists in an analysis of causation by means of regularities. Rather, as has been anticipated in the introduction and as will be substantiated in this chapter, the problems of an overly simple Humean account are resolvable by a suitable conceptual differentiation within the regularity framework.

The word “cause” is used very ambiguously in ordinary life and even in science. Sometimes it means a necessary, but it may be insufficient, condition. (e.g., “sparks cause fires”). Sometimes it means a sufficient, but it may be more than sufficient, condition or set of conditions (e.g., “Falling from a cliff causes concussion”). Sometimes it means a set of conditions which are severally necessary and jointly sufficient. But, in any interpretation, it involves one or both of the notions of “necessary” and “sufficient” condition.⁴

Broad here expresses the basic idea behind all modern regularity accounts of causation, as professed by e.g. Mackie (1974), May (1999), or Graßhoff and May (2001): General causation (or the notion of causal relevance) is to be analyzed by an appropriately combined recourse to the notions of a sufficient and of a necessary condition. Causes are sometimes sufficient, sometimes insufficient and sometimes severally necessary and jointly sufficient for their effects. Differentiating between the various cases and finding a definitionally appropriate combination of these notions is the business of a regularity account of causation.

Sufficient and necessary conditions are both well established logical notions that receive precise definitions in every introductory course in standard logic. Lemmon, for instance, defines:

(…) we shall say that, whenever it is the case that if P then Q , P is a *sufficient condition* for Q , and, whenever it is the case that only if P then Q , P is a *necessary condition* for Q .⁵

² Cf. Hobbes (1997), ch. IX, section 3.

³ Cf. Baumgartner and Graßhoff (2004), pp. 87-91.

⁴ Broad (1930), pp. 304-305.

⁵ Lemmon (1978 (1965)), p. 28.

Or expressed symbolically: P is a sufficient condition for Q iff (3.1) holds, while P is a necessary condition for Q iff (3.2) holds.

$$P \rightarrow Q \quad (3.1)$$

$$Q \rightarrow P \quad (3.2)$$

The relations “... is a sufficient condition for...” and “... is a necessary condition for...”, as (3.1) and (3.2) demonstrate, are converse relations: P is a sufficient condition for Q iff Q is a necessary condition for P .

Apart from the apparent inadequacy of a simple identification of type level causes with sufficient or necessary conditions and token level causes with the instances of their type level counterparts, some of the features of singular and general causation listed at the end of the previous chapter cast serious doubt on the analyzability of these relations by means of notions defined in propositional logic, as is the case for sufficient and necessary conditions defined along the lines of (3.1) and (3.2). For instance, consider the exclusions of self-causation (NSC) on token level and of conditional dependencies among factors of complex and alternative causes (CI), or take the fact that for a complex cause to become causally effective, its instances need to be coincidentally related (COIN): Neither of these features can be expressed within the formalism of propositional logic. (NSC) stipulates that all events that instantiate causing factors are different from the caused events. Formally expressing this constraint presupposes quantification over events and the identity predicate, both of which are not available in propositional logic. (CI), professing that neither factors in complex causes nor alternative causes include each other, similarly presumes quantification over events. Finally, (COIN) determines a special kind of relation to hold among the factors of a complex cause for this cause to become causally effective, while (PROX) stipulates spatiotemporal proximity for causes and effects – relations, of course, not being provided by propositional logic. This shows that formally accounting for the features of causal relevance uncovered in the previous chapter, calls for a formalism at least as rich as first-order predicate logic with identity.

Hence, before we can tackle the problem of putting sufficient and necessary conditions to a combined use as regards an analysis of causation, the relations “... is a sufficient condition for...” and “... is a necessary condition for...” need to be adequately expressed in first-order logic. The coming up three sections will be concerned with this translation. Moreover, in the course of these conceptual preliminaries a concept of *minimality* – essentially induced by (PR) – will be introduced and implemented to minimize sufficient and necessary conditions. This will yield the notion of a *minimal theory* which, in turn, will prove to be of crucial importance to the consecutive analysis of causal relevance.

3.2 Sufficient and Necessary Conditions in First-Order Logic

Prima facie, it might be thought that the conversion of (3.1) and (3.2) to the first-order formalism is straightforward: Just replace P and Q by any well-formed first-order formulas ϕ and ψ , such that the subjunctive remains the main operator, yielding ϕ to be a sufficient condition of ψ iff (3.3) holds and a necessary condition of ψ iff (3.4) holds:

$$\phi \rightarrow \psi \quad (3.3)$$

$$\psi \rightarrow \phi. \quad (3.4)$$

Yet, if the notions of a sufficient and a necessary condition were to be first-order defined in this vein, a causal interpretation of sufficient and necessary conditions would be precluded due to formal restraints. Cause events are not identical to effect events and, moreover, the former stand in a specific – yet to be introduced spatiotemporal relation – to the latter. Expressing these constraints by first-order means requires cause and effect variables to be contained within the same quantifier scope. However, any transformation of (3.1) and (3.2) into the first-order formalism along the lines of (3.3) and (3.4), such that the subjunctive remains the main operator, would prohibit cause and effect events from appearing within the same quantifier scope. We thus need a notion of first-order sufficient and necessary conditions that differs from (3.3) and (3.4).

Broad (1930) proposes the following first-order definitions:

“ C is a *sufficient condition* (...) of E ” means “Everything that has C has E ”.

“ C is a *necessary condition* (...) of E ” means “Everything that has E has C ”.⁶

In accordance with May (1999), I adopt Broad’s proposal and define ϕ to be a *sufficient condition* of ψ iff (3.5) holds, and ϕ to be a *necessary condition* of ψ iff (3.6) holds:

$$\forall \mu (\phi \mu \rightarrow \psi \mu) \quad (3.5)$$

$$\forall \mu (\psi \mu \rightarrow \phi \mu). \quad (3.6)$$

μ is to be read as a metavariable running over variables and $\phi \mu$ and $\psi \mu$ stand for any well-formed formulas with at least one free occurrence of μ . (3.6) is equivalent to (3.7), which for some does better at capturing the intuition behind the notion of a necessary condition. We will therefore make use of (3.7) every now and then when speaking of necessary conditions.

$$\forall \mu (\neg \phi \mu \rightarrow \neg \psi \mu) \quad (3.7)$$

⁶ Broad (1930), p. 306.

Sufficient and necessary conditions thus defined are, as in the case of (3.1) and (3.2), converse relations. ϕ is a sufficient condition of ψ iff ψ is a necessary condition of ϕ . That means, the following equivalence holds:

$$\forall\mu(\phi\mu \rightarrow \psi\mu) \dashv\vdash \forall\mu(\neg\psi\mu \rightarrow \neg\phi\mu) \quad (3.8)$$

It will be illustrative to look at some concrete examples of sufficient and necessary conditions. In light of (3.5), A being a sufficient condition of B can be stated by

$$\forall x(Ax \rightarrow Bx). \quad (3.9)$$

As shown in section 2.2.2, Ax is to be read as “ x instantiates A ”, which, in turn, is explicitly expressed within the set theoretic formalism by $x \in A$. (3.9) can thus be equivalently expressed by

$$\forall x(x \in A \rightarrow x \in B).$$

For reasons of brevity and transparency, we will, whenever convenient, stick to the symbolism of first-order predicate logic as in (3.9). Clearly though, translations to the set theoretic formalism are straightforward and always possible.

Based on (3.6) and (3.7), A being a necessary condition of B is equivalently formulated by (3.10) and (3.11).

$$\forall x(Bx \rightarrow Ax) \quad (3.10)$$

$$\forall x(\neg Ax \rightarrow \neg Bx) \quad (3.11)$$

(3.5) and (3.6) are open for more complex substitutions. (3.12), for instance, states that an event neither instantiating A nor B is a sufficient condition for there to be an event of type C that is related in terms of a relation W to the event that is neither A nor B .

$$\forall x(\neg(Ax \vee Bx) \rightarrow \exists y(Cy \wedge Wxy)) \quad (3.12)$$

(3.9) or (3.10) do not express what regularity theorists have in mind when they claim that type level causes should be analyzed in terms of sufficient and/or necessary conditions of their effects and token level causes in terms of instances of causally relevant factors. (3.9), for instance, maintains that whenever an event is of type A , this same event is of type B . This would, for example, be the case if A were to be interpreted as $\{x : x \text{ is a soccer game}\}$ and B as $\{x : x \text{ is a sport event}\}$. Of course, if thus interpreted, (3.9) expresses a dependency of set inclusion between A and B and not a causal relationship. In order for sufficient conditions to be causally interpretable, at least in principle, it must be guaranteed that antecedent and consequent of a material conditional are instantiated by *different* events.

That is not the only additional requirement the causal interpretability of sufficient conditions imposes. Recall, for instance, Hume’s definition of a cause as an event (type) that is always *followed* by another event (type). Hume argues in favor of a causal interpretation of conditionals of type “Whenever an instance of

the cause factor is given, there is an instance of the effect factor, which differs from the cause event and which succeeds the cause event". By requiring effects to succeed their causes, Hume builds the direction of time into his analysis of causation, thereby claiming the direction of time to be primary. Moreover, Hume excludes the possibility of simultaneous and backward causation by definition. All of these consequences are usually taken to be unwarranted.⁷ First, in section 3.6.2 it will be shown that an adequate analysis of the direction of causation is possible on purely logical grounds, such that the feasibility of an analysis of the direction of time by means of the direction of causation is maintained. Hence, since there is no theoretical need to integrate an external asymmetry or non-symmetry as the direction of time into an analysis of causation, conceptual parsimony and the prospect of a causal account of the direction of time command to abstain from such an integration. Second, whether there is simultaneous or backward causation clearly is an empirical and not a conceptual matter. If time travel is impossible, this impossibility is due to physical and not to conceptual constraints.

Nonetheless, Hume is right in requiring some spatiotemporal relation to hold between the instances of causes and effects. Causes do not determine their effects to occur anywhere and anytime, but close by, i.e. within a certain spatiotemporal frame or within the same situation (PROX). Yet, it is far from clear what exactly this spatiotemporal relation is. Depending on the causal process at hand, instances of causes and effects can be said to be properly related only if they are in direct spatiotemporal contact, while in other cases causes may well occur far away from their effects. The relation among causes and effects cannot be fixed to a specific spatiotemporal interval within which their instances are required to occur. The theory of Special Relativity only provides an upper bound for this interval: Causes must be instantiated within the past light cones of their effects. Notwithstanding this lacking specificity, given a concrete causal process it is normally uncontroversial which events can be said to be properly related in order to be amenable to a causal interpretation. If the striking of a match in Switzerland is followed by a match catching fire in England, it is clear that, even if the interval between these two events is thus that they are not excluded to be causally related by Special Relativity, the striking in Switzerland does not cause the light in England. On the other hand, given a concrete run of an experiment, say a number of substances are brought together in a test tube, it is commonly presumed that no events occurring in the course of this run are excluded from a causal dependency due to inadequate spatiotemporal relatedness. In this sense, causal dependencies are relativized to some *spatiotemporal frame*. The fuzziness involved in that notion reflects the impossibility to fix a specific interval for spatiotemporal proximity of causal relata.⁸ In chapter 5 it will be shown that depending on the process under investigation the spatiotemporal interval constituting a suitable frame for the involved factors can be

⁷ Cf. section 2.3.1.

⁸ Accordingly, whenever the spatiotemporal relation between causes and effects is explicitly discussed in theoretical accounts of causation, it is commonly left as unspecified as possible (cf. e.g. Xu (1997), p. 159-160.)

experimentally tested and adjusted. That means, even though the notion of a frame cannot be clarified and firmed up in an overall way, it can be spelled out relative to a given process. Until the experimental delineating of spatiotemporal frames will be discussed in chapter 5, we shall for now just assume that, given a concrete causal process, it is sufficiently clear what the spatiotemporal frame is within which the events involved in this process occur.

If sufficient and necessary conditions should, at least in principle, be open for causal interpretations, the syntax of acceptable substitutions in (3.5) and (3.6) must be restricted such that a token level cause x and a token level effect y are required to be two different events, i.e. $x \neq y$, that occur in the same spatiotemporal frame. In order to formally represent this spatiotemporal association, we introduce the relation Rxy , whose interpretation shall be fixed to e.g. "... occurs in the same spatiotemporal frame as ..." or "... occurs in the same situation as ...". R shall be taken to be symmetric, reflexive, and transitive. That means, R is an equivalence relation:

$$\begin{aligned} &\forall x\forall y(Rxy \rightarrow Ryx) \\ &\quad \forall x(Rxx) \\ &\forall x\forall y\forall z(Rxy \wedge Ryz \rightarrow Rxz). \end{aligned}$$

Contrary to Hume, who would have given R an asymmetric interpretation, our R thus does not involve any sort of asymmetry or non-symmetry. It is not to be interpreted in terms of a Humean "... occurs before ...". The reasons for this deviance from Hume essentially stem from the aforementioned fact that we shall account for the non-symmetry of general causation on purely logical grounds and, therefore, do not need to integrate an external non-symmetry into our analysis.⁹ Our theory of causation will not exclude simultaneous and backward causation on definitional grounds. Moreover, due to the notorious vagueness of R , we want our analysis of causation to be as independent of R as possible. R shall not do any analytical work for us. In this vein, the symmetry of R makes sure that the requirement to the effect that the instances of sufficient conditions need to be related in terms of R to their consequent does not covertly introduce causal notions into the notion of a sufficient condition. The symmetry of R makes sure that R cannot be misinterpreted as "... causes..." – for the latter is non-symmetric.

While the reflexiveness of R seems obvious in light of an interpretation along the lines of "... occurs in the same situation as ...", its transitivity might not seem quite so evident. Causal reasoning always operates within a given factor frame and analyzes the causal dependencies among the factors in that frame within a certain observational or experimental setting. The people of a given population are examined with respect to their suffering from lung cancer or the chemical mixture in a given test tube is studied as regards its explosiveness. Whatever happens within such a context is possibly causally related, at least a causal relation is not a priori

⁹ Cf. section 1.2.5.

excluded due to improper spatiotemporal relatedness. In this sense, R selects the setting in which a respective causal process takes place. If two pairs of events $\langle x, y \rangle$ and $\langle y, z \rangle$ occur in a certain setting, the pair $\langle x, z \rangle$ occurs in that setting, too. Apart from this, the transitivity of R is in line with a Humean interpretation in terms of "...occurs before ...". Furthermore, we shall see later on, that an intransitive or non-transitive R would be too strong in the sense that it would do substantial analytical work, that, as mentioned above, we do not want to cede to a relation lacking an interpretation that can be satisfactorily clarified.¹⁰

R is not fixed to be a binary relation. In complex causal structures R will have many more arguments. In case of three arguments, R can be taken to be "...and ... occur in the same spatiotemporal frame as ...".¹¹

Despite its fuzziness, R is indispensable for an accurate representation of causal relevance statements within first-order logic, for, without this relation, any instances of causally related factors would have to be seen as standing in the cause-effect relation. In this sense, R selects the instances of the effect factor that are amenable to be caused by a given cause event. R is needed in order to allow for spelling out singular causation based on a foregoing analysis of general causation. If it has been established that a factor A is causally relevant to a factor B , it is R which identifies the causes of each instance of B among the instances of A . Thus, an instance a of A being a cause of an instance b of B will be analyzed with recourse to the causal relevance of A to B in combination with $a \neq b$ and Rab .¹²

Equipped with $x \neq y$ and Rxy , $\psi\mu$ in (3.5) and (3.6) shall be restricted to formulas of the form

$$\exists y(Zy \wedge x \neq y \wedge Rxy), \quad (3.13)$$

where Z can be seen as any factor with y running over its instances. (3.13) allows for a formalization of what, in a Humean vein, could be claimed to be a causally interpretable material conditional:

$$\forall x(Ax \rightarrow \exists y(By \wedge x \neq y \wedge Rxy)). \quad (3.14)$$

(3.14) states that any event x being of type A is a sufficient condition for the existence of a second and different event of type B in the same spatiotemporal frame as x . As causal structures involve much more than two factors a thorough first-order analysis of causal relevance would require rather intricate formulas. For most practical purposes, however, a full first-order formalization of sufficient and necessary conditions is not needed. In order to conveniently abbreviate our notation we thus introduce " \mapsto ":

$$Z_1 \mapsto Z_2 =_{df} \forall x(Z_1x \rightarrow \exists y(Z_2y \wedge x \neq y \wedge Rxy)) \quad (3.15)$$

¹⁰ Cf. section 4.6. For further details on the interpretation of R see also 5.5.2.

¹¹ Strictly speaking, since R is an equivalence relation it could be confined to be binary. Three events e_1 , e_2 , and e_3 being related in terms of R follows from any two pairs of these events being R -related. Yet, such as not having to express R -relatedness in a pairwise manner, R shall not be fixed to be binary in the present context.

¹² Cf. section 3.7.

This abbreviated notation will be elaborated and extended in the following. It will be our main analytical instrument. Abbreviated expressions as $A \mapsto B$ can be read as ordinary material conditionals in propositional logic, as long as A and B are implicitly restricted to a specific kind of sentences. A and B state instantiations of two factors by *different* events that occur in the same *situation* or in the same *spatiotemporal frame*, such that, in view of a concrete factor frame, it is known what it means for those factors to be instantiated in a common situation. Even though this notational convention dispenses with a complete formalization of what is really meant by a purportedly causally interpretable regularity statement, it provides the central means for an analysis of causal relevance. Given that one speaks of events that instantiate the factors A and B and given that these events are different and occur in a certain spatiotemporal frame, “When A is instantiated, B is instantiated” is all a Humean regularity theorist proposes as analysans of causal relevance. However, such as to not forget the restriction of propositional conditionals to a very specific kind of sentences, we shall nonetheless every now and then – mostly in the course of exposing the conceptual fundament of our analysis – resort to first-order formalizations.

3.3 Minimally Sufficient Conditions

For various reasons that have been sketched in the introductory remarks to this chapter, Hume’s analysis of what it means for a factor to be a cause of another factor is unsatisfactory. One of the main reasons that have given rise to objections against Humean regularity accounts is the monotony of conditionals. A being sufficient for a factor B implies that the conjunction consisting of A and any other factor C is sufficient for B . Or formally:

$$\begin{aligned} \forall x(Ax \rightarrow \exists y(By \wedge x \neq y \wedge Rxy)) \vdash \\ \forall x(Ax \wedge Cx \rightarrow \exists y(By \wedge x \neq y \wedge Rxy)), \end{aligned} \quad (3.16)$$

which in our abbreviated symbolism corresponds to:

$$A \mapsto B \vdash A \wedge C \mapsto B. \quad (3.17)$$

Any true conditional stays true if any (true or false) conjunct is added to its antecedent. This, of course, seriously hampers a causal interpretation of e.g. (??). Suppose, A in fact is causally relevance to B . This by no means implies that the combination of A with *any* other factor is causally relevant to B as well. A causal interpretation of sufficient conditions is only warranted if the conditions exclusively contain factors that are essential to the bringing about of the purported effect. Redundant factors as C in (3.16) cannot be incorporated in causally interpretable sufficient conditions.

Virtually all modern regularity accounts of causation agree with respect to how this problem is to be solved: Sufficient conditions, in order to be open for causal

interpretations, need to be *minimalized*.¹³ To this end, we introduce the notion of a *minimally sufficient* condition:

Minimally sufficient condition: A conjunction of factors $A_1 \wedge A_2 \wedge \dots \wedge A_n$ is a minimally sufficient condition of a factor B iff

- (a) $A_1 \wedge A_2 \wedge \dots \wedge A_n \mapsto B$
- (b) there is no proper part α of $A_1 \wedge A_2 \wedge \dots \wedge A_n$ such that $\alpha \mapsto B$.

A “proper part” of a conjunction designates the result of any reduction of this conjunction by one conjunct. (b) can thus be more explicitly expressed as follows:

$$\neg(A_2 \wedge A_3 \wedge \dots \wedge A_n \mapsto B) \wedge \neg(A_1 \wedge A_3 \wedge \dots \wedge A_n \mapsto B) \wedge \dots \\ \dots \wedge \neg(A_1 \wedge \dots \wedge A_{n-2} \wedge A_n \mapsto B) \wedge \neg(A_1 \wedge \dots \wedge A_{n-2} \wedge A_{n-1} \mapsto B).$$

If a sufficient condition consists of one conjunct only, the proper part shall be defined to be the empty antecedent. However, an empty antecedent would only be sufficient for a universally instantiated factor. Since we have excluded predicates with all-embracing event extensions from factor definitions, there are no universally instantiated factors by definition. Hence, sufficient conditions consisting of one single conjunct are automatically minimally sufficient conditions.

In order to illustrate the theoretical work done by the notion of a minimally sufficient condition, consider $A \wedge B \wedge C$ to be a minimally sufficient condition for E . That means:

$$(A \wedge B \wedge C \mapsto E) \wedge \neg(A \wedge B \mapsto E) \wedge \neg(A \wedge C \mapsto E) \wedge \neg(B \wedge C \mapsto E). \quad (3.18)$$

First of all, note that $A \wedge B \wedge C$ reduced by *one* conjunct not being sufficient for E implies that $A \wedge B \wedge C$ reduced by *more* than one conjunct is not sufficient for E . For suppose e.g. $A \mapsto E$ were true. That would imply the truth of, say, $A \wedge B \mapsto E$, which contradicts (3.18). Furthermore, e.g. $\neg(A \wedge B \mapsto E)$ is true iff there is an event that instantiates A and B without there being a different event of type E which is related in terms of R to the instance of A and B , and analogously for $\neg(A \wedge C \mapsto E)$ and $\neg(B \wedge C \mapsto E)$. Hence, the notion of a minimally sufficient condition has important existential implications. If $A \wedge B \wedge C$ is minimally sufficient for E , there is at least one instance of *every* proper part of $A \wedge B \wedge C$ without a corresponding instance of E . More generally: For every conjunction γ_k with k conjuncts, where $k \geq 1$, and every conjunction γ_n with n conjuncts, where $n > 1$, such that $k < n$ and γ_k is the result of reducing γ_n by

¹³ Cf. Broad (1930), p. 306, Broad (1944), pp. 13-17, Mackie (1974), p. 62, May (1999), pp. 59-60, Graßhoff and May (2001), p. 94, Baumgartner and Graßhoff (2004), pp. 103-104, Hall (2004). Notwithstanding these standard logical minimalization procedures, regularity theories have been criticized for allegedly not being able to eliminate redundancies resulting from the monotony of sufficient (or necessary) conditions in non-causal – and thus non-circular – ways. In this vein, e.g. Brand and Swain (1970) do not take into account the purely non-causal criterion of what it means for a sufficient condition to be minimal proposed by Broad et al.

$n-k$ conjuncts: If γ_n is minimally sufficient for a factor Z then there is an instance of γ_k such that there is no corresponding instance of Z . For if this would not be the case, γ_n would not be minimally sufficient for Z .¹⁴

The notion of a minimally sufficient condition has yet another interesting implication. Given that $A \wedge B \wedge C$ is minimally sufficient for E , as expressed in (3.18), it follows that A , B , and C are conditionally independent. That is, if $A \wedge B \wedge C$ were to be interpreted as a complex cause of E , (C1) would be satisfied for that complex cause. For suppose A and B to be conditionally dependent, i.e. $A \rightarrow B$. Whenever A is instantiated, B is instantiated, too. In this case, $A \wedge C$ would be sufficient for E , and hence (3.18) would be violated, which contradicts the assumption above. The following is a theorem:

$$\begin{aligned} & \forall x(Ax \wedge Bx \wedge Cx \rightarrow \exists y(Ey \wedge x \neq y \wedge Rxy)) \wedge \\ & \forall x(Ax \rightarrow Bx) \vdash \forall x(Ax \wedge Cx \rightarrow \exists y(Ey \wedge x \neq y \wedge Rxy)). \end{aligned} \quad (3.19)$$

That means, two conditionally dependent factors – by definition – cannot be contained in the same minimally sufficient condition. Should minimally sufficient conditions thus be causally interpretable, the satisfaction of (C1) would be guaranteed.

However, minimally sufficient conditions as defined thus far are not straightforwardly causally interpretable, for no causes – regardless of their complexity – are ever going to be minimally sufficient in the sense defined above. Consider a match being struck in Bern, a match being dry in London, oxygen being present in Paris, and a match having a head of inflammable chemicals in Berlin. In this constellation, all of the causally relevant factors for a match to catch fire are instantiated, yet it does by no means follow that any of the involved matches actually light. As shown in section 2.3.4, the factors of a complex cause only become causally effective if *coincidentally* instantiated. Hence, if at all, causes can only be analyzed by means of the notion of a minimally sufficient condition, if the factors of a minimally sufficient condition are required to be coincidentally instantiated. In order to symbolically represent coincident instantiations, we introduce the n -ary relation K with n being the number of conjuncts in a minimally sufficient condition apart from K itself. K subsists among the instances of the factors in a minimally sufficient condition iff these factors are coincidentally instantiated. K can be seen on a par with any ordinary non-redundant factor within a minimally sufficient condition.¹⁵ If K does not hold among the instances of a conjunction of factors, the instantiations of these factors are not sufficient for the effect to occur. Thus, in contrast to R , K may remain uninterpreted. Of course, one manifest interpretation of K will be “... occurs in the same spatiotemporal frame as ...”, however, this is not the only possible interpretation of K . Take, for instance, the example of the Elm Street explosion on page 46 above. Relative to this context, the trigger switch and

¹⁴ Cf. May (1999), p. 62.

¹⁵ This will be slightly adapted for negative factors in section 3.6.4.

the detonator constituting a complex cause of the explosion can be said to be coincidentally instantiated if “... is connected by electrical conductors with ...” holds among their instances.¹⁶

Only the subset of minimally sufficient conditions that include K can possibly be causally interpreted. A possibly causally interpretable minimally sufficient condition is of the form:

$$\begin{aligned} \forall x_1 \forall x_2 \dots \forall x_n (A_1 x_1 \wedge A_2 x_2 \wedge \dots \wedge A_n x_n \wedge K x_1 x_2 \dots x_n \rightarrow \\ \exists y (B y \wedge x_1 \neq y \wedge x_2 \neq y \wedge \dots \wedge x_n \neq y \wedge R x_1 x_2 \dots x_n y)) \end{aligned} \quad (3.20)$$

As (3.20) demonstrates, the factors of a possibly causally interpretable minimally sufficient condition are not required to be instantiated by the same event, as has been implicitly presupposed thus far. As long as they occur coincidentally, factors in a complex cause may well be instantiated by different events.

The complexity of (3.20), which describes a minimally sufficient condition by explicitly mentioning three factors only, apparently calls for abbreviations. Our abbreviated notation has to be accommodated and extended. To this end we adopt the convention that a conjunction of factors whose instances are related in terms of K shall simply be concatenated without conjunctive and without explicit mention of K . A universally or existentially quantified conjunction of factors $A_1 x_1 \wedge A_2 x_2 \wedge \dots \wedge A_n x_n \wedge K x_1 x_2 \dots x_n$ will thus be represented by $A_1 A_2 \dots A_n$.

$$A_1 A_2 \dots A_n =_{df} A_1 x_1 \wedge A_2 x_2 \wedge \dots \wedge A_n x_n \wedge K x_1 x_2 \dots x_n \quad (3.21)$$

The quantifiers that bind the variables on the right-hand side of (3.21) can be left unspecified, because we will only use this abbreviated notation in connection with “ \mapsto ”, whose antecedent is determined to be universally quantified and whose consequent is existentially quantified by definition. Therefore, the context in which expressions of type $A_1 A_2 \dots A_n$ appear will always clarify the nature of the quantifiers involved. Given this notational convention, (3.20) can be transparently stated thus:

$$A_1 A_2 \dots A_n \mapsto B.$$

3.4 Minimally Necessary Conditions

Like sufficient conditions, necessary conditions may contain redundant factors. A being necessary for B implies that $A \vee C$ is necessary for B . Or formally:

$$\begin{aligned} \forall x (B x \rightarrow \exists y (A y \wedge x \neq y \wedge R x y)) \vdash \\ \forall x (B x \rightarrow \exists y ((A y \vee C y) \wedge x \neq y \wedge R x y)), \end{aligned} \quad (3.22)$$

which in our abbreviated symbolism corresponds to:

$$B \mapsto A \vdash B \mapsto A \vee C. \quad (3.23)$$

¹⁶ Further details concerning the interpretation of K are provided in section 5.5.2.

Any true conditional stays true if any (true or false) disjunct is added to its consequent. Analogous to the case of sufficient conditions, the extendability of necessary conditions by redundant disjuncts forecloses a causal interpretability of necessary conditions. Assume, A in fact is a causally interpretable necessary condition for B . This does not imply that $A \vee C$ can be considered a cause of B . Take A to be the presence of oxygen, which is causally relevant for a match to catch fire when struck (B). If there is no oxygen present, matches do not light. A is thus a necessary condition for B . Now, take C to be the singing of a song. $A \vee C$ is, although necessary for B , by no means causally interpretable. A causal interpretation of necessary conditions is only warranted if the conditions exclusively contain factors that are essential to the bringing about of the purported effect. Arbitrary factors as C in (3.22) cannot be incorporated in causally interpretable necessary conditions.

Broad (1944), May (1999), and Graßhoff and May (2001) have proposed an analogous solution to this problem as in the case of sufficient conditions. They call for a minimalization of necessary conditions.¹⁷ To this end, the notion of a *minimally necessary* condition is introduced:

Minimally necessary condition: A disjunction of factors $A_1 \vee A_2 \vee \dots \vee A_n$ is a minimally necessary condition of a factor B iff

- (a) $B \mapsto A_1 \vee A_2 \vee \dots \vee A_n$
- (b) there is no proper part β of $A_1 \vee A_2 \vee \dots \vee A_n$ such that $B \mapsto \beta$.

A “proper part” of a disjunction designates the result of any reduction of this disjunction by one disjunct. (b) can thus be more explicitly expressed as follows:

$$\neg(B \mapsto A_2 \vee A_3 \vee \dots \vee A_n) \wedge \neg(B \mapsto A_1 \vee A_3 \vee \dots \vee A_n) \wedge \dots \\ \dots \wedge \neg(B \mapsto A_1 \vee \dots \vee A_{n-2} \vee A_n) \wedge \neg(B \mapsto A_1 \vee \dots \vee A_{n-2} \vee A_{n-1}).$$

If a necessary condition consists of one disjunct only, the proper part shall be defined to be the empty consequent. However, an empty consequent would only be necessary for a factor without instances. Since we have excluded predicates with empty event extensions from factor definitions, there are no universally uninstantiated factors by definition. Hence, necessary conditions consisting of one single disjunct are automatically minimally necessary conditions.

As in the case of sufficient conditions, in order to illustrate some important implications of the notion of a minimally necessary condition, consider $A \vee B \vee C$ to be a minimally necessary condition for E . That means:

$$(E \mapsto A \vee B \vee C) \wedge \neg(E \mapsto A \vee B) \wedge \neg(E \mapsto A \vee C) \wedge \neg(E \mapsto B \vee C). \quad (3.24)$$

¹⁷ As shown by May (1999) and Graßhoff and May (2001), the fact that Mackie (1974) did not call for such a minimalization of necessary conditions was, in the end, responsible for the collapse of his regularity account based on so-called INUS-conditions. Not minimalizing necessary condition gave rise to the well-known Manchester-Hooters counterexample against Mackie’s theory. In Baumgartner and Graßhoff (2004) we have discussed the Manchester-Hooters in all detail along with a solution to this problem, that is rendered possible by minimalizing necessary conditions.

Parallel to what has been said about minimally sufficient conditions, note that $A \vee B \vee C$ reduced by *one* disjunct not being necessary for E implies that $A \vee B \vee C$ reduced by *more* than one disjunct is not necessary for E . For suppose e.g. $E \mapsto A$ were true. That would imply the truth of, say, $E \mapsto A \vee B$, which contradicts (3.24).

Furthermore, e.g. $\neg(E \mapsto A \vee B)$ is true iff there is an event that satisfies its antecedent without there being a corresponding event satisfying its consequent, and analogously for $\neg(E \mapsto A \vee C)$ and $\neg(E \mapsto B \vee C)$. Therefore, $A \vee B \vee C$ being minimally necessary for E implies that there is an instance of E without a corresponding instance of $A \vee B$, an instance of E without a corresponding instance of $A \vee C$, and an instance of E without a corresponding instance of $B \vee C$. More generally: For every disjunction δ_k with k disjuncts, where $k \geq 1$, and every disjunction δ_n with n disjuncts, where $n > 1$, such that $k < n$ and δ_k is the result of reducing δ_n by $n - k$ disjuncts: If δ_n is minimally necessary for a factor Z , then there is an instance of Z such that there is no corresponding instance of any disjunct in δ_k . For if this would not be the case, δ_n would not be minimally necessary. Now suppose, δ_n is

$$Z_1 \vee Z_2 \vee \dots \vee Z_i \vee \dots \vee Z_n,$$

and take δ_k to be

$$Z_1 \vee Z_2 \vee \dots \vee Z_{i-1} \vee Z_{i+1} \vee \dots \vee Z_n,$$

which results from δ_n by removing exactly one disjunct: Z_i , $1 \leq i \leq n$. We have already seen that the minimal necessity of δ_n implies that there is an event that instantiates Z without there being a corresponding instance of δ_k . Suppose this instance of Z to be event a . Now, either there is an instance of Z_i properly related to a or there is no such instance of Z_i . If there is no such instance, there would be an instance of Z without a corresponding instance of δ_n . This, however, contradicts the assumption that δ_n is necessary for Z . Therefore, there must be an instance of Z_i that is properly related to a . Hence, every disjunct of δ_n is instantiated at least once along with Z , when *all* the other disjuncts are absent. This proves that minimalizing necessary conditions – analogously to the minimalization of sufficient conditions – has very important existential implications. Every disjunct of a minimally necessary condition for a factor Z , is instantiated along with Z at least once when all the other disjuncts in that condition are absent.¹⁸

The notion of a minimally necessary condition has yet another interesting implication. Given that $A \vee B \vee C$ is minimally necessary for E , as expressed in (3.24), it follows that A , B , and C are conditionally independent. That is, if $A \vee B \vee C$ were to be interpreted as a disjunction of alternative causes of E , (C1) would be satisfied for these alternative causes. For suppose A and B to be conditionally dependent, i.e. $A \rightarrow B$. Whenever B is not instantiated, A is not instantiated either. In this case, $B \vee C$ would be necessary for E , and hence (3.24) would be violated,

¹⁸ Cf. May (1999), pp. 67-68.

which contradicts the assumption above. The following is a theorem:

$$\begin{aligned} \forall x(Ex \mapsto \exists y((Ay \wedge x \neq y \wedge Rxy) \vee (By \wedge x \neq y \wedge Rxy))) \\ \wedge \forall x(Ax \rightarrow Bx) \vdash \forall x(Ex \mapsto \exists y(By \wedge x \neq y \wedge Rxy)). \end{aligned} \quad (3.25)$$

That means, two conditionally dependent factors – by definition – cannot be contained in the same minimally necessary condition. Should minimally necessary conditions be causally interpretable, the satisfaction of (CI) would be guaranteed.

Yet, as in the case of minimally sufficient conditions, minimally necessary conditions are not causally interpretable – as alternative causes – without further specifications, for so far nothing has been said about the inner structure of the disjuncts in a minimally necessary condition. Thus far, they may well contain redundancies, and redundant factors, as the examples in this and the previous section have shown, are not amenable to a causal interpretation. Nonetheless, the path that leads to an elimination of remaining redundancies is well paved by now.

3.5 Minimal Theories

The yet lacking minimalization of the disjuncts in minimally necessary conditions is achieved by requiring that disjuncts in a minimally necessary condition be minimally sufficient conditions. This yields quantifier scopes in disjunctive normal form that are free of all redundancies. Reflecting this maximal minimalization, minimally necessary disjunctions of minimally sufficient conditions for a given factor shall be labelled *minimal theories* of the respective factor.¹⁹ Minimal theories will subsequently serve as central constituent of our analysis of causal relevance.

Minimal theory (I): A minimal theory of a factor B is a minimally necessary disjunction of minimally sufficient conditions of B , such that

- (a) conjuncts in each disjunct are required to be coincidentally instantiated,
- (b) B is required to be instantiated in the same spatiotemporal frame as its minimally sufficient conditions,
- (c) the instances of B differ from the instances of its minimally sufficient conditions.

Before minimal theories can be formally represented, the abbreviated notation initiated in the previous sections needs to be extended. Minimally sufficient and minimally necessary conditions have been defined as open conjunctions and open disjunctions, respectively. In order to account for that openness, two types of variables running over factors shall be introduced. For conjunctions of unknown or unspecified factors within sufficient conditions we shall implement the variables

¹⁹ Cf. Graßhoff and May (2001), May (1999), ch. 4, Baumgartner and Graßhoff (2004), ch. 5.

X_1, X_2 , etc. Thus, these X -variables are to be read as running over factors that are not explicitly integrated within sufficient conditions.

$$X_i =_{df} Z_1x_1 \wedge Z_2x_2 \wedge Z_3x_3 \wedge \dots \wedge Z_nx_n, n \geq 1,$$

with $i = 1, 2, 3, \dots$ and quantification depending on whether X_i appears right or left of “ \mapsto ”²⁰.

Building on the definition of X_i , we define the variables Y_A, Y_B , etc. to represent disjunctions whose disjuncts are not explicitly integrated within necessary conditions. The subscripts in case of the Y -variables correspond to the factors whose necessary condition a respective Y -variable complements.

$$Y_x =_{df} X_1 \vee X_2 \vee X_3 \vee \dots \vee X_n, n \geq 1,$$

with $x = A, B, C, \dots$ and quantification equally depending on whether Y_x appears right or left of “ \mapsto ”. We want to be able to operate with X - and Y - variables as with ordinary factors. To this end, some further notational conventions have to be introduced.

$$\begin{aligned} AX_i &=_{df} Ax_1 \wedge Z_1y_1 \wedge \dots \wedge Z_ny_n \wedge Kx_1y_1 \dots y_n, n \geq 1 \\ AY_x &=_{df} AX_1 \vee AX_2 \vee AX_3 \vee \dots \vee AX_n, n \geq 1 \\ \overline{X}_i &=_{df} \neg Z_1x_1 \vee \neg Z_2x_2 \vee \dots \vee \neg Z_nx_n \\ \overline{Y}_x &=_{df} \overline{X}_1 \wedge \overline{X}_2 \wedge \dots \wedge \overline{X}_n. \end{aligned}$$

With these notational means at hand, a factor A being part of a sufficient condition for B , such that this sufficient condition, in turn, is part of a necessary condition for B , can be expressed by (3.26).

$$(AX_1 \vee Y_B \mapsto B) \wedge (B \mapsto AX_1 \vee Y_B). \quad (3.26)$$

(3.26) states that whenever A is instantiated coincidentally with other factors X_1 , B is instantiated in the same spatiotemporal frame by an event that differs from the instances of AX_1 , and whenever B is instantiated, there is either a coincident instantiation of AX_1 or one of the disjuncts in the domain of Y_B is instantiated in the same spatiotemporal frame, such that the instances of B and of $AX_1 \vee Y_B$ differ. (3.26) can thus be seen as an abbreviation of expressions of the form of (3.27), where the incompleteness is indicated by dots instead of X - and Y -variables and k and i stand for arbitrary natural numbers.²¹

²⁰ Cf. page 98 above.

²¹ In (3.27) sufficient conditions are not disjunctively assembled as in (3.26). This divergence from (3.26) is warranted by the following logical equivalence that allows for a disjunctive assembling and a corresponding disintegration of sufficient conditions:

$$\begin{aligned} \forall x(Ax \rightarrow \exists y(Cy \wedge x \neq y \wedge Rxy)) \wedge \forall x(Bx \rightarrow \exists y(Cy \wedge x \neq y \wedge Rxy)) \dashv\vdash \\ \forall x(Ax \vee Bx \rightarrow \exists y(Cy \wedge x \neq y \wedge Rxy)). \end{aligned}$$

Or abbreviated:

$$(A \mapsto C) \wedge (B \mapsto C) \dashv\vdash A \vee B \mapsto C.$$

$$\begin{aligned}
& \forall x_1 \dots \forall x_i ((A_1 x_1 \wedge A_2 x_2 \wedge \dots \wedge A_i x_i \wedge K x_1 x_2 \dots x_i) \rightarrow \exists y (B y \wedge x_1 \neq y \wedge \dots \wedge x_i \neq y \wedge R x_1 \dots x_i y)) \wedge \\
& \dots \wedge \forall x_1 \dots \forall x_k ((Z_1 x_1 \wedge Z_2 x_2 \wedge \dots \wedge Z_k x_k \wedge K x_1 x_2 \dots x_k) \rightarrow \exists y (B y \wedge x_1 \neq y \wedge \dots \wedge x_k \neq y \wedge R x_1 \dots x_k y)) \wedge \\
& \forall y (B y \rightarrow (\exists x_1 \dots \exists x_i (A_1 x_1 \wedge A_2 x_2 \wedge \dots \wedge A_i x_i \wedge K x_1 x_2 \dots x_i \wedge x_1 \neq y \wedge \dots \wedge x_i \neq y \wedge R x_1 \dots x_i y) \vee \\
& \dots \vee \exists x_1 \dots \exists x_k (Z_1 x_1 \wedge Z_2 x_2 \wedge \dots \wedge Z_k x_k \wedge K x_1 x_2 \dots x_k \wedge x_1 \neq y \wedge \dots \wedge x_k \neq y \wedge R x_1 \dots x_k y)))
\end{aligned} \tag{3.27}$$

Emphasis needs to be put on the fact that (3.26) and (3.27) do not express a minimal theory of B . In order to turn (3.26) and (3.27) into a minimal theory, it has to be guaranteed that both sufficient and necessary conditions are minimal. A first and crucial step in this direction is done in (3.28).

$$\begin{aligned}
& (AX_1 \vee Y_B \mapsto B) \wedge \neg(A \mapsto B) \wedge \neg(X_1 \mapsto B) \wedge \\
& (B \mapsto AX_1 \vee Y_B) \wedge \neg(B \mapsto AX_1) \wedge \neg(B \mapsto Y_B).
\end{aligned} \tag{3.28}$$

(3.28) stipulates that A is non-redundantly contained in a minimally sufficient condition AX_1 of B such that AX_1 is non-redundantly contained in a minimally necessary condition of B . If it is furthermore granted that Y_B only consists of minimally sufficient conditions which are all part of a minimally necessary condition of B , we arrive at the called for complete minimalization of (3.26) and (3.27). Since expressions of type (3.28), whose Y_x is minimalized in this vein, will be of crucial importance for the subsequent discussions, we shall abbreviate them by expressions of type (3.29).

$$AX_1 \vee Y_B \Rightarrow B. \tag{3.29}$$

(3.29) will be termed a *double-conditional*.²² $AX_1 \vee Y_B$ is the antecedent of the double-conditional (3.29) and B its consequent. A double-conditional is a syntactic abbreviation for a conjunction of the form $(X_1 \vee Y_Z \mapsto Z) \wedge (Z \mapsto X_1 \vee Y_Z)$, such that its antecedent consists of a minimally necessary disjunction of minimally sufficient conditions for the consequent. This yields a second and more formal definition of the notion of a minimal theory.

Minimal theory (II): A minimal theory of a factor B is a double-conditional of the form $X_1 \vee Y_B \Rightarrow B$.

A factor Z is said to *be part of* a minimal theory of B iff Z is a conjunct of at least one disjunct in the antecedent of a minimal theory of B .

If expressions of type (3.29) were to be causally interpreted, many of the features of causal relevance uncovered in the previous chapter would be mirrored. Assume we take AX_1 to be a complex cause of B and Y_B to run over an open number of alternative causes of B . AX_1 being a sufficient condition of B satisfies the Principle of Determinism (PD). $AX_1 \vee Y_B$ being a necessary condition of B reflects the Principle of Causality (PC). Both sufficient and necessary conditions being minimalized satisfies the requirements of the Principle of Relevance (PR).

²² Cf. Baumgartner and Graßhoff (2004), p. 97.

For, as we have seen, minimalizing sufficient and necessary conditions guarantees for the absence of redundancies and, moreover, for the existence of at least one instance of AX_1 such that no other disjunct in Y_B is instantiated – and analogously for every disjunct in Y_B . In addition, the minimalization of sufficient and necessary conditions, in accordance with (CI), prohibits conditional dependencies among complex and alternative causes. Furthermore, the relations R and K , implicitly contained in (3.29), do justice to the constraints imposed by (PROX) and (COIN). Moreover, as we shall see in more detail in section 3.6.2 below, (3.29) allows for spelling out the direction of causation (NSY). In (3.29) AX and Y_B determine B to be instantiated, whereas B does only determine either AX or Y_B to be instantiated. B does not induce *which* of AX and Y_B is instantiated on a given occasion. By means of this non-symmetry of determination the non-symmetry of general causation can be accounted for. Finally, section 3.7 will show that a causal interpretation of expressions of type (3.29) even paves the way towards an adequate formal representation of the irreflexivity of singular causation (NSC).

Over and above, (3.29) reflects the complexity of causal structures. No factors are minimally sufficient for other factors in isolation. Causes are complexes of coincidentally instantiated factors and usually there are way over two alternative causes for each effect – the amount of alternative causes mostly being unknown. One way to account for the complexity of causal structures that include many unknown factors would be to prefix regularity statements as $A \mapsto B$ by a *ceteris-paribus* clause, as does e.g. Mackie (1974).²³ This would amount to relativizing regularity statements to a causal background or a causal field.²⁴ A could then be held to be minimally sufficient within a certain causal background, i.e. given coincident co-instantiations of an unknown number of other causally relevant factors. However, *ceteris-paribus* clauses are vague and, moreover, tend to lead into definitional circles. For determining what is to be considered the *causal* background of a given regularity structure obviously presupposes clarity as regards the notion of causal relevance.²⁵ (3.29) demonstrates that the notions of minimally sufficient and minimally necessary conditions, as defined in sections 3.3 and 3.4, are thus equipped as to account for the complexity of causal structures without such relativizations. (3.29) is open for arbitrary extensions of factor frames. The minimally sufficient condition AX_1 could, given appropriate evidence, be extended by explicitly integrating factors in the domain of X_1 into AX_1 . This, for instance, could yield that ACX'_1 is minimally sufficient for B , X_1 becoming X'_1 for its domain has changed – one factor represented by X_1 is not represented by X'_1 . Similarly, the minimally necessary disjunction of minimally sufficient conditions is extendable by explicitly integrating disjuncts in the domain of Y_B into $AX_1 \vee Y_B$. This, for instance, could yield that $ACX'_1 \vee DX_2 \vee Y'_B$ is a minimally necessary disjunction of minimally sufficient conditions of B . Criteria regulating extensions of expressions of

²³ We have argued in Baumgartner and Graßhoff (2004), ch. 5, that Mackie's implementation of a *ceteris-paribus* clause has considerable drawbacks.

²⁴ Cf. May (1999), ch. 5.

²⁵ Cf. section 1.2.2.

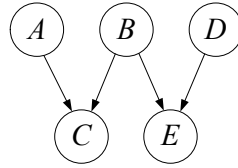


Fig. 3.1: A simple epiphenomenon that demonstrates the utility of (PPR).

type (3.29) will be discussed later. For now, it suffices to point out that (3.29) can account for the complexity of causal structures without relativizations to causal backgrounds or causal fields. Minimal theories are open for factor frame extensions and they constrain these extensions in a way that guarantees for the absence of redundancies.

Nonetheless, one of the features of causal relevance portrayed in the previous chapter is not mirrored by minimal theories: The Principle of Persistent Relevance (PPR). (PPR) requires causal relevancies to persist upon arbitrary extensions of factor frames, yet minimal theories, although being open for extensions, do not require factors to remain part of them if new factors are integrated into their antecedent. For illustrative purposes consider the example depicted in figure 3.1, which, by the way, can be seen as a variant of Mackie's (1974) famous example of the Manchester factory hooters.²⁶ Assume, for the sake of simplicity, that the causal graph in figure 3.1 exhibits the complete causal structure underlying the behavior of factors C and E . Hence, there are no unknown causal factors involved in the generation of events of type C and E .

There are three minimally sufficient conditions for E in this structure: \overline{AC} , B , D .²⁷ Now, suppose the scientific discipline investigating the causal structure behind events of type E first discovers that \overline{AC} and D are minimally sufficient for E . At the same time, the scientists concerned with E are confronted with instances of E in spatiotemporal frames in which both \overline{AC} and D are absent. We who know the real causal structure behind E , of course, can easily account for these cases: They are cases in which A , B , E , \overline{D} , and C are instantiated. At this stage of scientific knowledge, however, the corresponding discipline will conjecture the validity of the following minimal theory:

$$\overline{AC} \vee D \vee Y_E \Rightarrow E. \quad (3.30)$$

After a while of further investigation it is discovered that the formerly unknown factor B constitutes an additional minimally sufficient condition of E , i.e. the factor frame of the causal analysis at hand is extended. Moreover, now the scientists

²⁶ Cf. Mackie (1974), p. 83 et seq., Baumgartner and Graßhoff (2004), pp. 99-103.

²⁷ Section 3.6.4 will show that negative factors are not as straightforwardly integratable in minimal theories as this paragraph suggests. The problems as regards the membership of negative factors in minimally sufficient or necessary conditions, however, shall be neglected at this point. For now, negative factors are treated just as positive factors with respect to their integration in minimal theories.

can account for all instances of E . That means, whenever E is instantiated, there is an instance of \overline{AC} , B , or D in the corresponding spatiotemporal frame. Thus, a necessary condition of E has been discovered. This finding, of course, immediately raises the question as to whether this necessary condition is minimal. A straightforward disjunctive integration of B into the antecedent of (3.30) yields:

$$\overline{AC} \vee B \vee D. \quad (3.31)$$

Unfortunately though, (3.31) is not a minimally necessary condition of E . There is a proper part of this disjunction that is necessary for E : $B \vee D$. Whenever \overline{AC} is instantiated, B is so too: $\overline{AC} \rightarrow B$. On the other hand, it is not the case that whenever B is instantiated, there also is an instance of \overline{AC} . There are instances of B when all the other disjuncts in (3.31) are absent, yet there are no instances of \overline{AC} that satisfy this condition. In section 3.4 we have seen that in constellations of this type \overline{AC} cannot be part of a minimally necessary condition. \overline{AC} becomes redundant and accordingly drops out of a minimalized necessary condition. Hence, extensions of factor frames can directly affect the membership of a factor or condition in a minimal theory. The fact that \overline{AC} is part of (3.30) does not guarantee that \overline{AC} remains a constituent of every minimal theory resulting from extending the factor frame of (3.30). A minimal theory per se does not prohibit its constituents from dropping out. Therefore, despite all their causally interpretable features, minimal theories are not directly causally interpretable.

3.6 Causal Relevance Analyzed

3.6.1 A First Proposal

Notwithstanding the fact that causal relevancies cannot be directly read off minimal theories, the latter constitute the core of the conceptual inventory needed for a successful regularity theoretic analysis of causation. This section develops such an analysis by first identifying the causally interpretable minimal theories and then gradually specifying and adjusting our analysis of causal relevance.

In order to delineate the subset of causally interpretable minimal theories within the set of all minimal theories, the requirements imposed by the Principle of Persistent Relevance (PPR) have yet to be met. Such as to do justice to (PPR), a factor cannot simply be attributed causal relevance if it is part of a minimal theory, but only if it furthermore stays part of that minimal theory across *all* extensions of the corresponding factor frame. This not being the case for \overline{AC} of the exemplary graph in figure 3.1 precludes \overline{AC} from a causal interpretation – in accordance with the causal structure depicted in that graph. We have thus arrived at a first proposal for an analysis of causal relevance. The subset of causally interpretable minimal theories consists of those minimal theories whose antecedent only contains factors that are not rendered redundant by factor frame extensions. Having identified the causally interpretable minimal theories, direct causal relevance can be spelled

out in an forthright way. Dependent thereof indirect causal relevance is definable, which finally leads to an analysis of causal relevance in general.²⁸

Direct causal relevance (MT_d^*): A factor A is directly causally relevant for a factor B iff

- (a) A is part of a minimal theory Φ of B
- (b) A stays part of Φ across all extensions of its factor frame.

Indirect causal relevance (MT_i^*): A factor A is indirectly causally relevant for a factor B iff there is a sequence S of factors $Z_1, Z_2, \dots, Z_n, n \geq 3$, such that $A = Z_1, B = Z_n$, and for each $i, 1 \leq i < n$: Z_i is directly causally relevant for Z_{i+1} in terms of MT_d^* .

Causal relevance (MT^*): A factor A is causally relevant for a factor B iff A is directly causally relevant for B in terms of MT_d^* or indirectly causally relevant for B in terms of MT_i^* .

The central definition within this triad evidently is MT_d^* , as the other two each explicitly presuppose MT_d^* . Given the set of all direct causal relevance relationships within a certain factor frame, MT_i^* determines the corresponding indirect causal dependencies. In this sense, MT_i^* merely assembles complex causal structures from direct dependencies among single factors. MT_i^* is thus nothing but the transitive closure of MT_d^* . It needs to be stressed again that the identification of a factor as a direct or an indirect cause is relativized to the factor frame under investigation. A factor that is attributed direct relevance by MT_d^* within one factor frame is ascribed indirect relevance by MT_i^* within another frame. Causal relevance, finally, already in the previous chapter having been defined as the superordinate relation of direct and indirect relevance, is straightforwardly defined once the two subordinate relations have been accounted for.

We shall see in the following that this bottom-up strategy, even though it might intuitively seem appealing, generates some unwelcome ambiguities when it comes to building up complex causal structures from direct relevancies. Accordingly, these three definitions have been labelled a *first definitional proposal*, their provisional character being indicated by “*”. Nonetheless, the analytical strategy underlying these three definitions will remain unchanged. MT_d^* , MT_i^* , and MT^* lie at the heart of the analysis of causal relevance proposed in this study.

As shown in the previous section, condition (a) of MT_d^* accounts for most of the features of causal relevance displayed in chapter 2. Two features are not covered by (a): The transitivity of causal relevance (TR) and the Principle of Persistent Relevance (PPR). (TR) is mirrored in MT_i^* and so as to meet the requirements imposed by (PPR) MT_d^* is complemented by condition (b). In order to illustrate the utility and analytical power of (b), consider the causal structure depicted in figure

²⁸ Strictly speaking, the following definitions only analyze causal relevance for positive factors, i.e. positive effects. Causal relevance for negative factors will be accounted for in section 3.6.4 below.

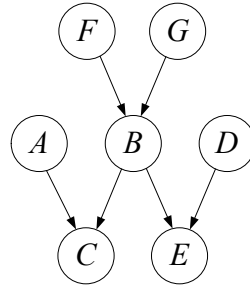


Fig. 3.2: The complete causal graph as regards the causes of B , C , and E .

3.2, which is a mere extension of the causal graph in figure 3.1. Again, for simplicity's sake, assume that the graph in 3.2 represents the complete causal structure regulating the behavior of B , C , and E , i.e. there are no further unknown causes of these three factors. This structure features two minimally sufficient conditions for B that are not causally interpretable: \overline{AC} and \overline{DE} . Analogously to the example of figure 3.1, suppose a causal investigation into the behavior of events of type B starts with discovering \overline{AC} and \overline{DE} to be minimally sufficient for B . These two conditions, however, do not account for all of the instances of B . Events of type B occur every once in while when neither of the two mentioned conditions is instantiated – for instance, when the as yet unknown factor F is instantiated along with A and D . This kind of evidence justifies conjecturing the validity of the following minimal theory:

$$\overline{AC} \vee \overline{DE} \vee Y_B \Rightarrow B. \quad (3.32)$$

Clearly, a causal interpretation of (3.32) would be utterly misleading. However, as in the example of the previous section, upon extensions of the factor frame by F and G , \overline{AC} and \overline{DE} will be deprived of their membership in (3.32), for neither of these conditions is part of a minimally necessary condition of B . Correspondingly, \overline{AC} and \overline{DE} , both of which are also minimally sufficient for F and G , will not remain part of the minimal theories of F and G , respectively. Extensions of the structure in figure 3.2 will sooner or later reveal factors – the causes of F and G – whose instances account for all the instances of F and G without recourse to \overline{AC} and \overline{DE} .

This finding is generalizable. Not only causes C_1, C_2, \dots, C_n are minimally sufficient for their effect E , the latter in combination with the absence of all its alternative causes save one is minimally sufficient for that remaining cause.

$$E\overline{C_2C_3} \dots \overline{C_n} \mapsto C_1 \quad (3.33)$$

However, these ‘causally backward’ minimally sufficient conditions do not remain part of minimal theories of pertaining factors across factor frame extensions, for they only account for a subset of the instances of C_1 – for the subset of instances that occur in combination with $\overline{C_2C_3} \dots \overline{C_n}$. This is a very specific subset of the

instances of C_1 . E cannot account for any event of type C_1 that occurs in combination with instances of any other factor C_i in $\{C_2, C_3, \dots, C_n\}$, such that C_1 and C_i overdetermine E or one of them preempts the other. As soon as the true causes of C_1 are taken into consideration, $\overline{EC_2C_3} \dots \overline{C_n}$ will become redundant and therefore not remain part of a minimally necessary condition of C_1 .

It might be remarked that if only R – which is implicitly contained in the consequent of expressions of type (3.33) – would be chosen to be some asymmetric relation as “...occurs before ...”, (3.33) would automatically be dispensed with, for E is not instantiated before C_1 . Plainly, if R is interpreted in this vein and backward causation is taken to be impossible on a priori grounds, causally backward minimally sufficient conditions would not be amenable to a causal interpretation in the first place. However, as the minimalization of necessary conditions is capable of dispensing with causally backward minimally sufficient conditions without such a recourse to the asymmetry of time, there is no theoretical need for fixing R to “...occurs before ...”. The direction of causation will be captured by means of the logical structure of minimal theories, i.e. without recourse to non-symmetries external to the conceptual framework of our analyses of causal relevance. Moreover, as we shall see shortly, building the direction of time into the analysis of causal relevance would have far-reaching theoretical consequences, that we want to bypass as long as we have a chance to do so.

MT_d^* thus excludes causally backward minimally sufficient conditions from a causal interpretation. Yet, how can it be known whether a minimally sufficient condition is causally backward before the corresponding factor frame has been fully extended? May (1999) has pointed to an important feature of causally backward conditions that allows for their identification prior to an exhaustive expansion of respective factor frames. Causally backward minimally sufficient conditions all contain causally dependent factors that *cannot be absent at the same time*, i.e. factors that are *subcontrary*. Given that $A \vee B$ stays part of a minimal theory of C across all factor frame extensions and is thus causally interpretable, the same cannot be the case for \overline{AC} with respect to a minimal theory of B . \overline{A} and C are causally dependent. Against the background of the causal graph in figure 3.2, these two factors cannot both be absent. If \overline{A} is not instantiated, i.e. if there is an event of type A , C cannot be absent, and if C is not instantiated, \overline{A} cannot be absent, i.e. there cannot be an event of type A . Hence, if factors within a minimally sufficient condition cannot be absent at the same time, i.e. if they are subcontrary, the corresponding minimally sufficient condition will not remain part of a minimal theory across all factor frame extensions. Causal dependencies among factors of minimally sufficient conditions are ‘minimalized away’ by MT_d^* as respective factor frames are extended.²⁹

Nonetheless, causally backward regularities as $\overline{AC} \mapsto B$ often are of crucial importance for causal investigations. They justify diagnostic inferences. Observing an effect to take place while knowing that certain alternative causes are not

²⁹ Cf. May (1999), p. 74.

instantiated on a respective occasion allows for diagnoses that account for the observed effect with recourse to its remaining causes. Representing causal structures by minimal theories hence not only allows for an identification of the causally interpretable regularities in a given factor frame, but, over and above, also accounts for diagnostic inferences.

As these examples show, minimal sufficiency is not sufficient for a factor to be a cause of another factor. Neither is minimal necessity. Take e.g. factor C in the graph of figure 3.2. C is causally backward minimally necessary for A , B , F , or G , since there is no proper part of C that is necessary for any of these factors. However, C is not sufficient for any of these factors. A mere instantiation of C does not determine which of the other factors is present on a given occasion. Therefore, according to MT_d^* , the minimally necessary condition consisting of C is not causally interpretable.

The introduction of condition (b) into MT_d^* might raise objections as to the practical applicability of a thus analyzed notion of causal relevance. (b) confines the applicability of that notion to factors that remain part of minimal theories across all possible extensions of respective factor frames, yet actual causal analyses always suffer from practical limitations as regards the complete expansion of investigated factor frames. In concrete experimental contexts it is impossible to ever fully expand investigated factor frames. Over the set of all events an unmanageable abundance of factors may be defined. Practical constraints preclude testing all these factors for a possible integration into a minimal theory. Therefore, it might be argued, irrespective of its conceptual adequacy the notion of direct causal relevance as defined by means of MT_d^* cannot ever be attributed to any factor whatsoever.

First of all, it needs to be stressed that this objection is of epistemic or practical nature only. It does not vitiate the conceptual expediency of MT_d^* . Whether a factor remains part of a given minimal theory across all factor frame extensions is an objective matter of fact, regardless of the practical testability of this property. Clearly though, our analysis of causal relevance shall not only be conceptually accurate, but applicable as well. Causal analyses, regardless of their theoretical background, always have to be conducted within very limited factor frames. These limitations do not normally impair methodologies of causal reasoning. Indeed, the applicability of MT_d^* is not jeopardized by condition (b). As customary for methodologies of causal reasoning, so-called *homogeneity* constraints, imposed on causally analyzed test situations, allow for determining whether a factor remains part of a minimal theory across all factor frame extensions even when only a very restricted area of the complete factor frame is actually investigated.³⁰ Chapter 5 will develop such a procedure of causal reasoning that implements MT_d^* and does not call for analyzing complete factor frames.

³⁰ Cf. Baumgartner and Graßhoff (2004), ch. 9.

3.6.2 The Direction of General Causation

General causation is *non-symmetric*: A being a causally relevant to B does not imply B being causally relevant to A nor does it imply \bar{B} being causally relevant to \bar{A} . Nevertheless, it is by no means excluded that A is causally relevant to B and B is causally relevant to A at the same time. On type level, there are causal cycles. For instance, with increasing unemployment the consumption of the population is reduced. This causes decreased profits on the side of the employers, which, in turn, causes them to lay off even more people. Thus, the unemployment increases anew. A regularity theory that takes general causation to be its primary analysandum, evidently, is required to account for the non-symmetry of general causation. This, however, is often claimed only to be possible on regularity theoretic grounds if recourse is made to e.g. the direction of time, i.e. to a non-symmetry which is external to the conceptual framework of MT_d^* . Yet, this section is going to substantiate what has already been anticipated in section 1.2.5, namely that, as there is no need for such a recourse in order to prevent MT_d^* from a causal interpretation of causally backward minimally sufficient or minimally necessary conditions, minimal theories, in contrast to first appearances, account for the non-symmetry of general causation without recourse to the direction of time. Furthermore, in section 3.7, we shall see that the non-symmetry of general causation induced by MT_d^* allows for a straightforward handling of the non-symmetry of singular causation.

At first sight, it might seem that double-conditionals are symmetrical. Consider anew factor B of the graph in figure 3.2. The causal generation of its instances is accounted for by the following minimal theory:

$$F \vee G \Rightarrow B. \quad (3.34)$$

Whenever $F \vee G$ is instantiated, so is B , and whenever there is an instance of B , there is one of $F \vee G$. Hence, the dependency between effects and their causes seems to be perfectly symmetrical. Undoubtedly, minimal theories are symmetrical with respect to their antecedent *as a whole* and their consequent. Nonetheless, the relation between *particular* disjuncts in the antecedent of a minimal theory and its consequent is not symmetrical. The instantiation of a particular disjunct is minimally sufficient for the consequent, but not vice versa. The instantiation of the consequent does not determine a particular disjunct to be instantiated.³¹ The consequent only determines the whole disjunction of minimally sufficient conditions. F and G in (3.34) are each minimally sufficient for B , the latter however is only minimally sufficient for $F \vee G$. B does neither determine F nor G to be instantiated. Hence, given that an instantiation of F is observed, it can be inferred that there is an event of type B somewhere in the corresponding spatiotemporal frame. On the other hand, if an event of type B is observed, no such inference to a proximate instantiation of F is possible. The observed instance of B might well have been caused by an event of type G . This non-symmetry corresponds to the non-symmetry of determination.

³¹ Cf. Graßhoff and May (2001), pp. 97-99.

However, in section 1.2.5 we have already seen that this analysis would not go through, if B were not to have *two* alternative causes. Suppose there to be only one minimally sufficient condition for B : F . If and only if F is instantiated, an event of type B occurs. Hence F is both minimally sufficient and minimally necessary for B , where, of course, emphasis needs to be put on the fact that the instances of F and B differ – otherwise, according to (FI), F and B would be identical.³² Given such a structure, B is minimally sufficient and minimally necessary for F as well. F and B are either co-instantiated or neither of the two factors is instantiated. It is impossible to instantiate one of the factors without the other. In such a constellation MT_d^* does not allow for deciding whether F is the cause of B or B is the cause of F . Against the background of MT_d^* , one has to abstain from a causal judgment. Is this consequence a merit or a defect of MT_d^* ? I know of no plausible example of a concrete causal process that would be structured in this simple way: $(F \mapsto B) \wedge (B \mapsto F)$. It therefore cannot be decided whether we would have any causal intuitions whatsoever if confronted with such a dependency relationship between two *different* factors. In lack of concrete examples, I take abstinence from a causal judgement as induced by MT_d^* to be a reasonable handling of two factors being mutually minimally sufficient and minimally necessary for each other.

It might be objected that, if we were to find that F is always instantiated *before* B , there would be perfectly reasonable grounds on which to identify F as the cause and B as the effect. Undoubtedly, time could be implemented as a basis that imposes a direction on the completely parallel behavior of F and B .³³ However, this would be a high price to pay. One would have to account for the direction of time independently of the direction of causation and thus deviate from Reichenbach's programme that takes the direction of causation to be primary.³⁴ For causal relevance then would have to be analyzed somehow along the following lines: A factor A is causally relevant for a factor B iff A is part of a minimally necessary disjunction (with number of disjuncts ≥ 1) of minimally sufficient conditions for B and instances of A occur before instances of B . Yet, due to the factual inexistence of such simplistic causal structures as described above, taking the direction of time to be primary is a futile theoretical maneuver. Clearly, the literature on causation is full of simplistic structures. Still, accounting for all the anomalous examples in the literature shall not be our goal here. Our aim is to account for actual causal dependencies in nature. In accordance with all theories of causation that set out to analyze the causal relation without recourse to an external non-symmetry, we therefore abstain from a causal interpretation of dependency structures as $(F \mapsto B) \wedge (B \mapsto F)$.³⁵ Symmetrical dependency relationships among factors are not causally interpretable, or to accentuate it further: Symmetrical dependency relationships among factors are not causal.

³² Cf. page 61.

³³ As is well known, many authors have gone this way. Most notably Hume (1999 (1748)), p. 146. Similarly Suppes (1970), p. 23.

³⁴ Cf. p. 2.3.1.

³⁵ Cf. Sanford (1976), Papineau (1989), Hausman (1986), Hausman (1998).

Manifestly, our analysis of the direction of general causation has an important implication as regards the minimal complexity of a causal structure. A conjunction of factors AX_1 that is both minimally sufficient and minimally necessary for another factor B cannot be identified as cause of B , for B would be minimally sufficient and minimally necessary for AX_1 as well. All empirical evidence such as a dependency structure generates consists in perfectly correlated instantiations of AX_1 and B – both being either co-instantiated or absent. As already mentioned, I contend that, in order to distinguish causes from effects and to orient the cause-effect relation, at least *two* alternative causes are needed for each effect.

Minimal complexity of a causal structure: The minimal complexity of a causal structure consists of two alternative causes for a given effect: $A \vee B \Rightarrow C$. $A \Rightarrow C$ is not causally interpretable.

Causation is not a *one-to-one*, but a *many-to-one* relation. There is a plurality of causes to each effect. Of course, an effect may also have multiple effects of its own, yet a factor does not cause one of its effects on one occasion and another on a different occasion.³⁶ There is no such thing as alternative effects. An effect, however, can be caused by different causes on different occasions.

Against our claim that minimal theories allow for capturing the non-symmetry of general causation without presupposing the direction of time it might be insisted that the whole theory of causation proposed in the present study presupposes the direction of time for its analysis of the notion of an event. In chapter 2 events have been defined to be particulars in time and space, which are assigned one single temporal locality by a function τ .³⁷ However, assigning a temporal locality to an event is independent of determining the direction of time. What is presupposed by τ is not the direction of time, but merely a *temporal order*. Temporal order can be pictured as a linear ordering that can be mapped onto a straight line. τ assigns one point on that line to every event. Yet, whether time ‘flows’ from left to right or from right to left is irrelevant to such an assignment. Or as Reichenbach puts it:

The discussion of the problem of time has greatly suffered from the confusion of two concepts, from neglecting the distinction between *order* and *direction*.

The points on a straight line, which is infinite on both sides, are arranged in a certain order; but the line does not possess a direction. When the points are in a linear order, or serial order, they are governed by an asymmetrical and transitive relation. (...) It is well known from the theory of relations that every asymmetrical, connected, and transitive relation establishes a serial order. When we say that a line, though serially ordered, does not have a direction, we mean that there is no way of distinguishing structurally between left and right, between the relation and its converse.³⁸

³⁶ Cf. Mackie (1974), p. 167.

³⁷ Cf. section 2.2.2.

³⁸ Reichenbach (1956), p. 26.

In order to determine which direction shall be labelled “left” or “right”, an additional regulation is called for. That means, the localizability of events does not presuppose the direction of time. Ordering events is one thing, deciding which of two events in an ordered pair occurs *before* the other is a different matter.

The direction of time is not the only external non-symmetry that has been proposed as a means to orient the generic cause-effect relation. Gasking (1955), von Wright (1971), and recently Price (1992) have argued that the direction of general causation is best accounted for with recourse to agency, manipulability, or intervention. Causes are those factors whose instances are manipulable. One can arbitrarily intervene on causes such that they are instantiated or suppressed, whereas the behavior of effects is determined by the behavior of their causes exclusively. Effects can only be instantiated or suppressed by mediation of their (directly manipulable) causes. This account has several drawbacks. For one, it relativizes the non-symmetry of causal relevance to human intervention, where intuitively this non-symmetry seems to be perfectly independent of human existence. Causal processes – as planetary movements or volcanic eruptions on Saturn – that are not manipulable by humans are non-symmetric and orientable just as the breaking of a window or the starting of a car engine. Moreover, it is unclear how the notions of agency, intervention, and manipulation could be clarified without recourse to causation. In fact, these notions seem to straight-out presuppose causation. Agency, intervention, and manipulation are nothing but a specific variant of causal processes – the variant that involves human action.

This is a generalizable consequence of implementing any external non-symmetry for a theoretical account of the direction of causal relevance: The external non-symmetry becomes more basic than the cause-effect relation. Thereby a straightforward causal analysis of these external non-symmetries is blocked. However, intervention, for instance, can hardly be more transparently analyzed than in terms of causal processes where human action is involved as a cause. As long as we are not inevitably constrained to an analysis of the direction of causation by means of an external non-symmetry, theoretical prudence calls for abstinence from recourse to such non-symmetries.

The advantages of an analysis of the direction of general causation by means of the non-symmetry of determination are perspicuous. First of all, it allows for conceptual parsimony. The core analyses – minimal theories – of causal relevance establishes a non-symmetry between generic causes and effects such that no recourse to external non-symmetries is called for. Second, this furnishes us with a notion of causal relevance that does not presuppose the direction of time or of human intervention, which, in turn, leaves these external non-symmetries amenable to a straightforward causal analysis – as e.g. induced by the Reichenbachian programme. Finally, by accounting for the generic cause-effect non-symmetry without resorting to the relation “... occurs before ...” we abstain from determining on a priori grounds that causes precede their effects. Whether there is backward causation or not, is a question to be answered by physical sciences a posteriori. If causes were taken to precede their effects by definition, the pertinent reaction to somebody

wanting to build a time machine would be to invoke that building a time machine is excluded on conceptual grounds. However, whether time machines are possible or not clearly seems to be a physical, not a conceptual matter. Our analysis of the cause-effect non-symmetry is in line with this intuition. Furthermore, we remain completely non-committal as regards controversial questions as to the possibility of simultaneous or backward causation.

3.6.3 Multiple Minimal Theories – Specification of Causal Structures

A factor can have more than one minimal theory that regulates its behavior. This, as we shall see in the next chapter, is not unproblematic, for not necessarily all of those multiple minimal theories are causally interpretable. Minimalizations of necessary conditions are not always free of ambiguities. These problems, however, shall be postponed for the moment. Notwithstanding the ambiguities, many of the multiple minimal theories of a factor are causally interpretable. There is thus more than one minimal theory that can be seen as representing the causal relevancies behind the behavior of a factor. This provides us with the analytical means to capture variable specifications of causal structures. Consider the following factors:

$$\begin{aligned} A &= \{x : x \text{ is a smoking of tobacco}\} \\ B &= \{x : x \text{ is a smoking of a cigarette}\} \\ C &= \{x : x \text{ is a smoking of a cigar/cigarillo}\} \\ D &= \{x : x \text{ is a smoking of a pipe}\} \\ E &= \{x : x \text{ is exposure to radioactivity}\} \\ L &= \{x : x \text{ is a suffering from lung cancer}\} \end{aligned}$$

A , on the one hand, and B , C , and D , on the other, cannot be part of the same minimal theory, as they are conditionally dependent and the Principle of Relevance determines that alternative minimally sufficient conditions as well as factors contained in one minimally sufficient condition are conditionally independent.³⁹ If it is furthermore assumed that tobacco can either be smoked by means of cigarettes, cigars, cigarillos, or pipes (and in no other way), the following set theoretic dependencies hold:

$$B \cup C \cup D = A, \quad \text{where } B \cap C = \emptyset, B \cap D = \emptyset, C \cap D = \emptyset.$$

Nonetheless, none of these factors is excluded from being attributed causal relevance for lung cancer by MT_d^* . A is part of one minimal theory, and B , C , D , since they are mutually conditionally independent, are part of another.

$$EX_1 \vee AX_2 \vee Y_L \Rightarrow L \quad (3.35)$$

$$EX_1 \vee BX_2 \vee CX_3 \vee DX_4 \vee Y_L \Rightarrow L. \quad (3.36)$$

³⁹ Cf. sections 3.3 and 3.4 above.

Predicates that define conditionally dependent factors describe the events that satisfy them on different levels of specification. Accordingly, (3.35) and (3.36) represent the causal structure underlying events of type L on different levels of specification. The X -variables in (3.35) and (3.36) are to be read as independent variables. Hence X_2 is not implied to represent the same conjunction of unknown or unspecified factors in (3.35) and (3.36). It is, however, well possible that X_2 in (3.35) stands for the same conjunction of factors as do X_2 , X_3 , and X_4 in (3.36).

Analogously to A , E could be specified as well:

$$\begin{aligned} F &= \{x : x \text{ is exposure to alpha radioactivity}\} \\ G &= \{x : x \text{ is exposure to beta radioactivity}\} \\ H &= \{x : x \text{ is exposure to gamma radioactivity}\} \end{aligned}$$

Again, none of these factors can be integrated either in (3.35) or in (3.36), for their instances are included in E . They are part of yet a third minimal theory of L :

$$FX_1 \vee GX_2 \vee HX_3 \vee BX_4 \vee CX_5 \vee DX_6 \vee Y_L \Rightarrow L, \quad (3.37)$$

where X -variables are again independent of the corresponding X -variables in (3.36).

Evidently, (3.35), (3.36), and (3.37) are not independent minimal theories. They can be said to be *materially equivalent* in the sense that for any event e : e either instantiates the antecedent of *all* three minimal theories or of *none* of them. The transition from A to B , C , D and from E to F , G , H , respectively, shall be referred to as *specification by supplementation*. This label captures the basic idea behind the mode of specification at hand: Factors are specified by conjunctively or adverbially *supplementing* a respective factor defining predicate. The thus resulting fine-grained predicate is satisfied by a proper subset of the events that satisfy the corresponding coarse-grained predicate. No minimal theory can both include a factor Z_1 and factors Z_2 and Z_3 that result from a specification by supplementation of Z_1 . Yet, if Z_1 is part of a minimal theory Φ , there is a materially equivalent minimal theory Ψ such that Z_2 and Z_3 are part of Ψ . This is illustrated by the following theorems:

$$\begin{aligned} \forall x(Ax \vee Bx \rightarrow \exists y(Gy \wedge x \neq y \wedge Rxy)) \wedge \forall x(Ax \equiv Cx \vee Dx) \vdash \\ \forall x(Cx \vee Dx \vee Bx \rightarrow \exists y(Gy \wedge x \neq y \wedge Rxy)) \end{aligned} \quad (3.38)$$

$$\begin{aligned} \forall y(Gy \rightarrow \exists x((Ax \vee Bx) \wedge x \neq y \wedge Rxy)) \wedge \forall x(Ax \equiv Cx \vee Dx) \vdash \\ \forall y(Gy \rightarrow \exists x((Cx \vee Dx \vee Bx) \wedge x \neq y \wedge Rxy)) \end{aligned} \quad (3.39)$$

There is another, however less straightforward, form of specifying a factor besides specification by supplementation. Consider the following factors:

$$\begin{aligned}
A &= \{x : x \text{ is an operating refrigerator}\} \\
B &= \{x : x \text{ is a vaporizing of ammonia by a compressor in a refrigerator}\} \\
C &= \{x : x \text{ is a pumping of vaporized ammonia gas through the coils of a} \\
&\quad \text{refrigerator}\} \\
D &= \{x : x \text{ is a cooling of food}\}
\end{aligned}$$

Essentially, an operating refrigerator does nothing else than pumping ammonia through its coils and vaporizing it in the walls of the cooling compartments such that the vaporized ammonia absorbs the heat from the compartments. Hence, the instances of A are causally structured events or processes. This causal structure can be dissected by analyzing it as a coincidence of B and C . As A can be considered causally relevant for D , thus can the coincidence BC . In this sense, B and C constitute a complex cause of D .

The transition from A to B and C is another form of specification. However, in this case A is not specified by conjunctively or adverbially supplementing its defining predicate, but by subdividing it in its two main *causal* components. Reflecting this causal subdivision we shall in this case speak of *causal specification*. For the sake of simplicity, let us assume that A is instantiated if and only if B and C are coincidentally instantiated. Yet, and that is a crucial difference between specification by supplementation and causal specification, BC is not required to be instantiated by the same events as A . The instances of, say, B are not operating refrigerators, as all instances of “smoking a cigarette” are instances of “smoking”. Still, if and only if there is an event of type A in the spatiotemporal frame of an instance of D , there also is an event of type B properly related to that instance of D . This does not follow from set theoretic (conditional) dependencies as in the previous example, but from causal dependencies, i.e. from the fact that BC is a causally specified description of the instances of A . Events of type B are a causal component of events of type A . Hence, for any events x : If x instantiates A , there is a coincident instantiation of BC in the spatiotemporal frame of x , and for any events y and z : If y instantiates B and z instantiates C and y and z are coincident, there is an instance of A in the spatiotemporal frame of y and z .

$$\begin{aligned}
\forall x(Ax \rightarrow \exists y\exists z(By \wedge Cz \wedge Kyz \wedge Rxyz)) \wedge \\
\forall y\forall z(By \wedge Cz \wedge Kyz \rightarrow \exists x(Ax \wedge Rxyz))
\end{aligned} \tag{3.40}$$

Note that (3.40) is not a causally interpretable dependency, for the instances of A , B , and C , though not necessarily identical, are not required to be different. Therefore, B and C are not causally relevant for A , rather and somewhat loosely spoken, they constitute A .

Can two causally specified factors as B and C be part of the same minimal theory as the coarse-grained A ? A and BC cannot be part of the same minimally sufficient condition, for eliminating either A or BC from a sufficient condition of D containing ABC does not result in a loss of sufficiency, since A guarantees for an instantiation of BC and vice versa. A sufficient condition $ABCX_1$ of D contains

at least two sufficient proper parts: AX_1 and BCX_1 . A and BC , therefore, cannot be part of the same minimally sufficient condition of D . Neither can A and BC be part of the same minimally necessary condition, for, as section 3.4 has shown, minimalizing a necessary condition δ for a factor Z implies that each disjunct in δ is instantiated at least once along with Z such that all the other disjuncts in δ are absent. This would obviously be violated by a disjunction that includes $A \vee BC$. These two disjuncts are either co-instantiated or both absent. Since A and BC can neither be part of the same minimally sufficient nor of the same minimally necessary condition, A and BC cannot be part of the same minimal theory. As in the case of specification by supplementation, in order to attribute causal relevance to causally specified factors, different minimal theories are called for:

$$AX_1 \vee Y_D \Rightarrow D \quad (3.41)$$

$$BCX_1 \vee Y_D \Rightarrow D. \quad (3.42)$$

There are some important differences between specification by supplementation and causal specification. Whereas any factor can be specified by supplementation, i.e. by adverbially or conjunctively adding further constraints to its defining predicates, provided only that conditions (a) to (d)⁴⁰ for factor defining predicates are satisfied, causal specification is much less straightforward. Whether a factor is causally specifiable depends on the causal structure of its instances. Moreover, whether there is a minimal theory which contains fine-grained causally specified factors cannot be deduced from the mere existence of a minimal theory containing the corresponding coarse-grained factor. There are no theorems as (3.38) and (3.39) that would guarantee for the existence of such fine-grained minimal theories. Whether there is a minimal theory that contains causally specified factors needs to be answered by empirical testing.

As we have seen in section 2.2.1 (p. 45), progressively uncovering causal structures is a very common and effective method for causal analysis. Progressive localization of causes starts with the analysis of causal structures among coarse-grained factors and then, by gradually specifying those factors, advances to uncover fine-grained structures. This section has shown that MT* is capable of attributing causal relevance to factors on different levels of specification. Minimal theories provide the means to represent causal structures on variable levels of specification.

3.6.4 *Minimal Theories and Negative Factors*

There is an extensive debate in the literature concerning the existence of negative relata of the causal relation, both on type and on token level.⁴¹ Intuitions differ as to whether absences are to count as real causes or effects. Did the father not preventing the child from running onto the street cause the accident, or are children not running onto the street causally relevant for absences of accidents?⁴² Some,

⁴⁰ Cf. section 2.2.2, p. 58.

⁴¹ Cf. e.g. Mackie (1992), Bennett (1987), Ellis (1990), Armstrong (1997), Dowe (2001).

⁴² Cf. Dowe (2001), p. 217-218.

as e.g. Ellis (1990) or Armstrong (1999), hold that a causal relation can only subsist between positives, whether factors or events. Others, as e.g. Broad (1930), Mackie (1974), or Schaffer (2000), while not setting forth anything like negative events as relata of singular causation either, accept both positive and negative factors as relata of general causation.⁴³

It is undisputed that there are general causal statements that refer to negatives both in cause and effect position.

(54) Not drinking alcohol is causally relevant for the absence of a headache.

(55) The absence of traffic light signalings is causally relevant for car crashes.

(56) Signalings of traffic lights are causally relevant for the absence of accidents.

Prima facie, (54), (55), and (56) are all acceptable general causal statements. Yet, does this prima facie acceptability stand the test of theoretical scrutiny or, as Armstrong (1999) puts it, is it the case that “ordinary *language* of causality conceals the true ontological situation”⁴⁴? Both denying and reserving negative factors and absences the status of possible relata of general causation can be argumentatively backed. There are two main types of arguments against the causal interpretability of absences: First, allowing absences to be causes or effects conflicts with the intuition that the instances of generic causes and effects to be entities in time and space, for, as we have seen in section 2.2.1, there is nothing in time and space that can be seen as the absence of a headache, of traffic lights, or of accidents. Second, causally interpreting absences gives rise to a counterintuitive proliferation of causes. A famous illustration of this argument is provided by Hart and Honoré (1985 (1959)):

A gardener whose duty it is to water flowers fails to do so and in consequence they die. It can be said that it is impossible to treat the gardener’s omission here as the cause unless we are prepared to say that the ‘failure’ on the part of everybody else to water the flowers was equally the cause (...).⁴⁵

Whereas it seems legitimate to interpret the absence of a watering by the gardener as the cause of the death of the flowers, we are not prepared to identify the Pope’s or Vladimir Putin’s failure to water the flowers as causes of their withering. Similarly, while the absence of traffic light signalings is an admissible cause of accidents at intersections, the absence of meteors or of earthquakes destroying the respective intersection before the crashing of two cars are normally not put forward as causes of accidents. Defenders of absences as causal relata usually

⁴³ Cf. Broad (1930), I, p. 311, Mackie (1974), ch. 3. Dowe (2001), p. 218, occupies a kind of intermediate position: He claims that, while absences are not relata of genuine causation, they are relata of what he calls *quasi-causation*. For the reason why negative events, for most, are unwelcome candidates for causal relata see section 2.2.1 above.

⁴⁴ Armstrong (1999), p. 177.

⁴⁵ Hart and Honoré (1985 (1959)), p. 38.

counter this argument from the proliferation of causes by pointing out that not only causally interpreted absences can be counterintuitively proliferated, the same holds for ordinary positive factors. Customary, accidents are not taken to be caused by Big-Bang or by the invention of the wheel. Hart and Honoré (1985 (1959)), for instance, argue that the problem at hand, rather than being a characteristic feature of causally interpreted absences, is a problem of distinguishing between causes that are of interest relative to a given causal inquiry and mere standing-conditions of the process under investigation.⁴⁶ This distinction can only be drawn on pragmatic grounds, which are not at issue in the course of an analysis of the notion of causal relevance. Notwithstanding the fact that the Pope would not be held responsible for the death of the flowers, his failure to water the flowers still was a cause of their untimely decay. Counterfactually spoken: Had he watered the flowers, they would not have died. The flowers faded away due to an absence of water, regardless of the person supplying the water. Had anyone watered the flowers, they would not have died.

Friends of the causal interpretability of absences and negative factors usually back up their position by advancing concrete examples which indispensably involve absences or negative factors as causes and effects. Schaffer (2000), for instance, describes the causal intermediaries between the piercing of a human heart and brain death as follows:

(...) the relevant intermediaries are *absences*: the heart piercing causes an absence of oxygenated blood traveling from the right ventricle, through the relevant arteries, to the brain, which absence causes an absence of oxygen resupply to the brain cells, which absence causes oxygen starvation.⁴⁷

It is not anything contained in the blood that causes brain death, but the absence of oxygen. The actual composition of the blood does not matter. Brain death is caused by something missing from the blood. Similarly for accidents that are caused by missing traffic lights or infections that are caused by missing vaccination.

Authors that oppose the causal interpretability of absences try to rebut examples of this kind by positively reformulating them. In some cases this works fine: A statement as (57) is easily reformulated in terms of (58).

- (57) The absence of water on a match is causally relevant for the match catching fire when struck.
- (58) The dryness of the match is causally relevant for the match catching fire when struck.

Such reformulations, however, are only possible in case of pairs of contradictory factor defining predicates as *wetness* and *dryness* such that the negation of one constituent of the pair extensionally coincides with the other. Whenever negations of factor defining predicates are not co-extensional with a positive factor defining

⁴⁶ Cf. Hart and Honoré (1985 (1959)), p. 38. Similarly Schaffer (2000), p. 295.

⁴⁷ Schaffer (2000), p. 294.

predicate, corresponding negative factors are not identical to positive factors. There are negative factors – as the absence of vaccination or the absence of oxygen in the blood – that need to be causally interpreted because they are not identical to any positive factors. Thus, even if the causal relevance of negative factors might in some cases be analyzable in terms of the relevance of identical positive factors, the causal interpretability of negative factors cannot generally be dismissed.

The section at hand shall therefore be concerned with the integration of negative factors into minimal theories and, hence, with the causal interpretation of negative factors within the framework of MT^* . The *prima facie* acceptability of (54), (55), and (56) only stands the test of theoretical scrutiny if the causal structures reported by those statements can be given analyses within a respective theory of causation, such that absences and negative factors are rendered amenable to causal interpretations without thereby implying that absences are instantiated in time and space. This and the next section shall show that MT^* provides the means to do justice to both of these requirements and thus is capable of causally interpreting absences. However, we shall see that an integration of negative factors into minimal theories calls for some further specifications of the notion of a minimal theory.

Negative Factors as Effects

(54), (55), and (56) illustrate that colloquial talk implements absences both in cause and effect position of causal statements. Thus far we have only been concerned with minimal theories with positive factors in effect position, i.e. with positive effects. Corresponding to every such minimal theory there exists a logically equivalent formula with a negative expression in the consequent of the first and the antecedent of the second conjunct. (3.45) and (3.46) illustrate this. While (3.45) corresponds to the syntax devised for minimal theories in section 3.5, the logically equivalent (3.46) does not.⁴⁸ By means of arrow and quantifier definitions every expression of the form of a minimal theory is transferrable into an equivalent expression of type (3.46). The details of this procedure are elucidated in the appendix on page 264.

$$\begin{aligned} \forall x_1 \forall x_2 ((Ax_1 \wedge Bx_2 \wedge Kx_1x_2) \vee (Cx_1 \wedge Dx_2 \wedge Kx_1x_2) \rightarrow \exists y (Ey \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \wedge \\ \forall y (Ey \rightarrow \exists x_1 \exists x_2 (((Ax_1 \wedge Bx_2 \wedge Kx_1x_2) \vee (Cx_1 \wedge Dx_2 \wedge Kx_1x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \end{aligned} \quad (3.45)$$

⁴⁸ The logical equivalence of (3.45) and (3.46) can be illustrated by means of propositional contraposition. By abstracting from spatiotemporal constraints and from the exclusion of self-causation (3.43) can be seen as the minimal theory in (3.45), expressed by means of propositional logic.

$$(A \wedge B) \vee (C \wedge D) \leftrightarrow E. \quad (3.43)$$

By contraposition ($P \leftrightarrow Q \dashv\vdash \neg P \leftrightarrow \neg Q$) (3.43) is equivalent to

$$\neg((A \wedge B) \vee (C \wedge D)) \leftrightarrow \neg E, \quad (3.44)$$

which corresponds to (3.46). A detailed first-order proof of the logical equivalence of (3.45) and (3.46) is provided in the appendix (p. 264).

$$\begin{aligned} & \forall y(\forall x_1\forall x_2\neg(((Ax_1\wedge Bx_2\wedge Kx_1x_2)\vee(Cx_1\wedge Dx_2\wedge Kx_1x_2))\wedge x_1\neq y\wedge x_2\neq y\wedge Rx_1x_2y)\rightarrow\neg Ey)\wedge \\ & \forall x_1\forall x_2(\forall y\neg(Ey\wedge x_1\neq y\wedge x_2\neq y\wedge Rx_1x_2y)\rightarrow\neg(((Ax_1\wedge Bx_2\wedge Kx_1x_2)\vee(Cx_1\wedge Dx_2\wedge Kx_1x_2)))) \end{aligned} \quad (3.46)$$

From the logical equivalence of (3.45) and (3.46) it follows that (3.45) represents a causal structure – as claimed by MT^* – if and only if (3.46) does so too. Both expressions state exactly the same, they thus explicate the same causal structure if and only if one of them explicates a causal structure. In order to see that (3.45) and (3.46) in fact represent the same causal structure, let us first verbalize the two formulas. As shown in section 3.5, (59) can be seen as a verbalization of a minimal theory as (3.45), while (3.46) could be verbalized along the lines of (60).

- (59) For all events x_1 and x_2 , if they coincidentally instantiate AB or CD , there is a different event y of type E that occurs in the same spatiotemporal frame as x_1 and x_2 ; and for all events y of type E , there are two events x_1 and x_2 that coincidentally instantiate AB or CD and that both differ from y and occur in the same spatiotemporal frame as y .
- (60) For every event y , if there are neither coincident instantiations of AB nor of CD in y 's spatiotemporal frame which differ from y , then y is of type \bar{E} ; and for all events x_1 and x_2 , if it is not the case that there is an event y of type E in the spatiotemporal frame of x_1 and x_2 which differs from x_1 and x_2 , then x_1 and x_2 are neither coincident instantiations of AB nor of CD .

For the purpose of illustrating that (3.45) and (3.46) exhibit the same causal structure, let us look at the example in (54). Consider the following factors:

$$\begin{aligned} A &= \{x : x \text{ is a drinking of alcohol}\} \\ B &= \{x : x \text{ is a smoking}\} \\ C &= \{x : x \text{ is a staring into a computer screen}\} \\ D &= \{x : x \text{ is an anxiety}\} \\ E &= \{x : x \text{ is a tension-type headache}\} \\ K &= \{\langle x, y \rangle : x \text{ and } y \text{ occur simultaneously to the same person}\} \end{aligned}$$

For simplicity's sake, let us assume that people suffer from tension-type headaches if and only if they drink and smoke or stare into a computer screen while being anxious. Hence, AB and CD are the two complex causes of E . By means of identifying AB and CD as the two minimally sufficient conditions of E that are jointly minimally necessary for E , (3.45) represents the causal structure at hand. These dependency relationships have a number of implications, as for instance, that people that do not drink and are not anxious do not suffer from tension-type headaches. The same holds e.g. for non-smokers that do not stare into computer screens. All in all, $AB \vee CD$ being a minimally necessary condition consisting of minimally sufficient conditions of E implies that there is no tension-type headache within a spatiotemporal frame if and only if there is neither a drinking and a smoking nor an

anxious staring into a computer screen within the respective frame. This is, loosely spoken, what (3.46) says. The sufficient and necessary condition for not suffering from tension-type headaches identified by (3.46) is satisfied iff neither A nor C , or neither A nor D , or neither B nor C , or neither B nor D are instantiated within a given spatiotemporal frame. This, in turn, implies that $AB \vee CD$ is a minimally necessary condition consisting of minimally sufficient conditions of E .

Hence, if an analysis of causal relevance in the vein of MT^* should turn out to be successful, (3.45) represents a causal structure if and only if (3.46) represents the same structure. By stating the causes of E one states the causes of \bar{E} and vice versa. It does not matter whether a causal structure is represented by identifying the causes of a positive effect or of its negation. Introducing minimal theories as double-conditional expressions of type (3.45) has been a mere syntactical convention. Minimal theories might just as well have been defined as expressions of type (3.46). This, of course, raises the question as to why the double-conditional form of (3.45) was chosen as the standard form of minimal theories. The reason for this choice is straightforward: syntactical transparency. (3.46) does not render minimally necessary and sufficient conditions of \bar{E} as transparent as does (3.45) in the case of the minimally sufficient conditions of E . The double-conditional syntax of minimal theories exhibits minimally sufficient conditions as *disjuncts* of a minimally necessary condition. These disjuncts are directly translatable into alternative complex causes of corresponding effects. A comparable transparency is not provided by (3.46). The minimally sufficient conditions of \bar{E} are \bar{AC} , \bar{AD} , \bar{BC} , and \bar{BD} . These conditions, however, are merely deducible from (3.46) and not directly displayed therein.

The logical equivalence of (3.45) and (3.46) shows that in order to account for the causes of negative factors, the standard double-conditional form of minimal theories does not need to be adapted. For, by identifying the causes of a positive effect, the causes of its negation are identified as well.⁴⁹ By holding A to be causally relevant for E , \bar{A} is stated to be causally relevant for \bar{E} . Causal relevance for negative factors is analyzable by means of causal relevance for positive factors, and hence by means of minimal theories as devised so far.

Causal relevance for negative factors (MT_n^*):⁵⁰ A factor Z (\bar{Z}) is causally relevant for a factor \bar{B} iff the negation of Z (\bar{Z}) is causally relevant for B according to MT^* .

MT_n^* determines \bar{A} to be causally relevant for \bar{B} iff MT^* identifies A to be causally relevant for B . Analogously, A can be said to be causally relevant for \bar{B} iff \bar{A} is designated to be causally relevant for B by MT^* . In order for a negative factor \bar{A} to be causally relevant for B , \bar{A} either needs to be part of a minimal theory of B or part of a sequence of factors in the sense of MT_i^* , such that B is the last element

⁴⁹ A similar idea has been sketched in Ragin (1987), pp. 98-99.

⁵⁰ The provisional character of MT^* is carried over to MT_n^* . Once MT will have received its terminal form, MT_n will be restated accordingly (cf. section 3.6.6, p. 137).

of the sequence. In either case, \overline{A} has to be integratable into minimal theories, i.e. into expressions of type (3.45). The next subsection will be concerned with such an integration.

Negative Factors as Causes

Recall our definitions of positive and negative factors in section 2.2.2. A positive factor A is the set of events in the extension of a predicate that complies to the conditions – (a) to (d) – imposed on factor defining predicates, whereas a negative factor \overline{A} is merely the complementary set of A . Now, consider the following factors:

$$\begin{aligned}
 A &= \{x : x \text{ is a simultaneous passing of cars}\} \\
 B &= \{x : x \text{ is a signalling of traffic lights}\} \\
 C &= \{x : x \text{ is drunken driving across an intersection}\} \\
 D &= \{x : x \text{ is a crashing of cars}\} \\
 K &= \{\langle x, y \rangle : x \text{ and } y \text{ occur simultaneously at the same intersection}\}
 \end{aligned} \tag{3.47}$$

For simplicity's sake, let us assume that crashes at intersections happen if and only if there either are cars simultaneously passing the intersection while there are no traffic lights regulating the traffic or drunken drivers are simultaneously passing the intersection. Given such a constellation, we want to say that $A\overline{B}$ and AC are the two (sufficient) complex causes of D . If negative factors are simply taken to be ordinary conjuncts within minimally sufficient conditions, (3.48) would have to be seen as the minimal theory adequately describing the causal context at hand.

$$\begin{aligned}
 \forall x_1 \forall x_2 ((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \wedge \\
 \forall y (Dy \rightarrow \exists x_1 \exists x_2 (((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y))
 \end{aligned} \tag{3.48}$$

However, relative to the above interpretations of the factors in (3.48), this cannot be a minimal theory, for factors A and \overline{B} are conditionally dependent:

$$\forall x (Ax \rightarrow \neg Bx). \tag{3.49}$$

Every simultaneous passing of cars is a non-signalling of traffic lights. As (3.50) and (3.51) below illustrate, whenever K is reflexive just as R , \overline{B} is redundant in the first disjunct of (3.48). Relative to a reflexive K , every instance of A is not only an instance of \overline{B} , but, moreover, is related in terms of K to itself. That means, the antecedent of (3.48) is satisfied merely given an instance of A . In view of (3.49), (3.48) does not require \overline{B} to be instantiated over and above an instantiation of A . (3.48), accordingly, is not a minimal theory after all. Thus, \overline{B} cannot be causally interpreted by MT^* , which contradicts the true causal structure assumed to underlie the behavior of the factors in the current example.

This argument has the form of a reductio. In order to evaluate which of its premisses needs to be adjusted such as to render our analysis of causal relevance

capable of adequately representing the causal structure at hand, let us review the involved premisses and assumptions.

- (I) K is reflexive.
- (II) Variables in minimal theories run over events.
- (III) $A =_{df} \{x : Ax\}$, where A is a proper factor defining predicate and x runs over events.
- (IV) $\bar{A} =_{df} \{x : \neg Ax\}$, where A is a proper factor defining predicate and x runs over events.
- (V) If $\forall x(Ax \rightarrow Bx) \wedge \neg \forall x(Bx \rightarrow Ax)$, B is redundant in a minimally sufficient condition containing A . If $\forall x(Ax \rightarrow \neg Bx) \wedge \neg \forall x(\neg Bx \rightarrow Ax)$, \bar{B} is redundant in a minimally sufficient condition containing A .⁵¹
- (VI) The analysis of causal relevance by means of minimal theories is correct.
- (VII) $A\bar{B}$ and AC are the two complete complex causes of D .

(VII) is unquestionable, for it merely describes the causal structure assumed for the example at hand. As (VI) constitutes the core of the analysis of causal relevance proposed in the present study, it apparently is immune against being abandoned as long as any of the other assumptions is not solidly secured. Therefore, before modifications of (VI) shall be taken into account let us concentrate on assumptions (I) to (V).

The argument against the integration of negative factors in minimal theories would not go through if K was not, as in (3.47), interpreted in terms of a reflexive relation. Despite its conditional dependence on A , \bar{B} does not become redundant within a sufficient condition for D that includes both A and a non-reflexive K , for (3.50) holds.⁵²

$$\begin{aligned} & (\forall x_1 \forall x_2 ((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \wedge \\ & \forall x Rxxx \wedge \forall x (Ax \rightarrow \neg Bx)) \not\vdash \forall x_1 (Ax_1 \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge Rx_1x_1y)). \end{aligned} \quad (3.50)$$

Yet, whenever K is given a reflexive interpretation, \bar{B} drops out of a sufficient condition for D containing A . (3.51) is a theorem.⁵³

$$\begin{aligned} & (\forall x_1 \forall x_2 ((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \wedge \forall x Rxxx \\ & \wedge \forall x Kxx \wedge \forall x (Ax \rightarrow \neg Bx)) \vdash \forall x_1 (Ax_1 \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge Rx_1x_1y)). \end{aligned} \quad (3.51)$$

This holds generally: Conditionally dependent factors do not become redundant in sufficient conditions if the interpretation of K is fixed to a non-reflexive relation.

⁵¹ Cf. section 3.3.

⁵² A falsifying instance for the negation of (3.50) can be found in the appendix (p. 264).

⁵³ A proof of (3.51) is given in the appendix (p. 265).

Yet, in section 3.3 K has been introduced as an ordinary relational factor on a par with unary factors, such that its interpretation, unlike R 's, is open. Therefore, a reflexive interpretation of K is not excluded by definition. The problems as regards the integration of negative factors in minimal theories now raise the question as to whether, contrary to primary intentions, K should be treated on a par with R and be furnished with a fixed interpretation. Of course, contrary to R , K would not have to be fixed to a reflexive, but to a non-reflexive relation. However, the forthright interpretations of K – as e.g. the one given in (3.47) or spatiotemporal contiguity or occurrence within the same experimental setting – all are reflexive relations. Moreover, leaving the interpretation of K open enables us to treat K as an ordinary non-redundant factor within minimally sufficient conditions. Diverging from such a straightforward treatment of K would be costly, for then, more would have to be said about possible interpretations of K – undoubtedly a very intricate task. The interpretation of K shall therefore be left open. Reflexive interpretations as in (3.47) are not only possible, but commonplace. Assumption (I) is perfectly acceptable.

On the basis of (II), (III), and (IV) A and \bar{B} , given the interpretations in (3.47), are conditionally dependent: Every simultaneous passing of cars is a non-signalling of traffic lights. Or differently: No event consisting in a signalling of traffic lights is an event that consists in cars simultaneously passing an intersection. This dependency stems from the fact that A and B are distinct factors, i.e. they have an empty intersection: $A \cap B = \emptyset$. MT^* , on the other hand, only attributes causal relevance to factors if they are and stay part of a minimal theory of a respective effect. The only double-conditional expression – as defined so far – that can be seen as representing the causal structure stipulated in (VII) is (3.48). Hence, (VI) and (VII) imply that (3.48) is a minimal theory of D . In contrast, (V), given the factor definitions in (3.47), determines \bar{B} to be redundant in (3.48), which means that (3.48) cannot be a minimal theory of D . All of this proves that assumptions (II) to (VII) are inconsistent.

In order to decide which of these assumptions needs to be given up or adapted, let us examine them in turn. It might not be apparent what role (II) plays in the above argument. The fact that factors are taken to be classes of events and that events are the objects of quantification is responsible for the conditional dependency of A and \bar{B} . There are no events that instantiate both A and B . Suppose however, we were not to quantify over events, but, say, spatiotemporal intervals, i.e. localities. Let us assume moreover, a spatiotemporal locality is said to instantiate a factor Z_1 if the property expressed by the predicate in the definiens of Z_1 is exemplified within the corresponding spatiotemporal interval. Relative to this theoretical framework, (3.49) does not hold any more, for there are localities in which A is instantiated while \bar{B} is absent, e.g. when traffic lights are regulating the simultaneous passing of cars. Hence, if variables in minimal theories were not taken to run over events but spatiotemporal localities, \bar{B} would not be rendered redundant by (V). This, in turn, would grant (3.48) the status of a minimal theory, which

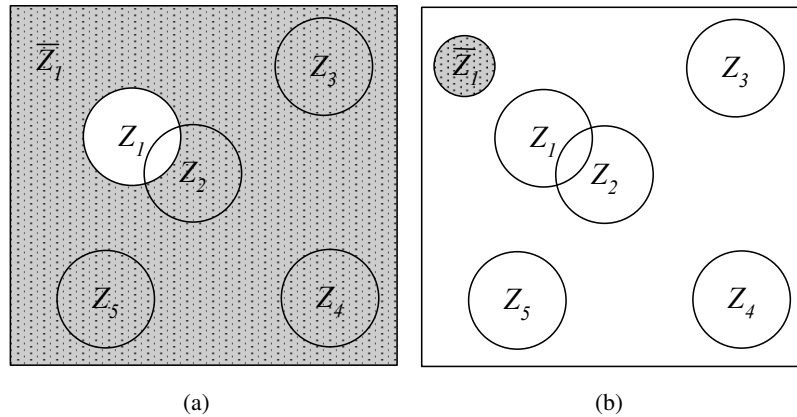


Fig. 3.3: $\overline{Z_1}$ relative to two different analyses of negative factors. In diagram (a) negative factors are taken to be the complementary set of corresponding positive factors, in diagram (b) negative factors are seen as sets of events that are independent of the instances of Z_1 , e.g. sets of negative events. In both cases, the box contains all events and the shaded area marks the instances of $\overline{Z_1}$.

would enable an adequate capturing of the causal structure behind the example at hand along the lines of MT^* .

However, quantification over spatiotemporal localities would induce an abandonment of event causation. An expression as Fa , where a is not an event, but a place in time and space, states a *fact*. Quantifying over space-time points would thus prompt a shift towards fact causation, which would render our theory of causation liable to all objections against fact causation discussed in chapter 2. As shown in section 2.2.1, none of these objections are conclusive such that giving up (II) would be outright impossible. Yet, taking the causal relata to be facts would have very broad philosophical implications and would require adaptations in several philosophical areas not inherently connected to causation. Resolving the problem of integrating negative factors into minimal theories by such a consequential departure from event causation only seems warranted if no other solution happens to be available. Before we decide on the faith of (II), let us therefore look at the other assumptions, while, for now, sticking to quantification over events.

What about our analysis of positive and negative factors Z_1 and $\overline{Z_1}$ as complementary sets of events? One consequence of this analysis is that whenever Z_1 or $\overline{Z_1}$ comprise only a few instances, the complementary set encompasses almost all events and thereby an abundance of other factors. This is illustrated in diagram (a) of figure 3.3. The shaded area marks the instances of $\overline{Z_1}$, which contains all factors (Z_3, Z_4, Z_5) that have an empty intersection with Z_1 . Thus, $\overline{Z_1}$ is conditionally dependent on $Z_3, Z_4,$ and Z_5 in the sense of (V). In the factor frame of diagram (a) $\overline{Z_1}$ can only be part of a minimal theory that contains Z_2 , for this is the only factor in that frame $\overline{Z_1}$ is not conditionally dependent on. From a purely formal point of view this consequence might not seem unacceptable. However, the example in

(3.47) clearly shows that there are causal structures that call for an integration of negative factors within a complex cause along with other factors on which they conditionally depend, according to the factor definitions in (III) and (IV). That means, regardless of conditional dependencies induced by (III) and (IV), positive and negative factors can be part of the same complex cause, which MT* mirrors by an integration of these factors in the same minimally sufficient condition. Yet, such a combined integration of positive and negative factors in minimally sufficient conditions, according to (V), is only possible if the corresponding factors are not conditionally dependent.

The conditional dependence of \overline{Z}_1 on Z_3 , Z_4 , and Z_5 could be broken by a modification of (IV). It might be held that a negative factor \overline{Z}_1 is not simply the complementary set of a corresponding positive factor Z_1 , but an event set that is independent of Z_1 , as, for instance, a set of *negative events*. Thus, it might be argued that spacetime is not only populated by positive, but also by negative events, such that the latter constitute the instances of negative factors. Diagram (b) of figure 3.3 illustrates the consequence of such an account as regards the conditional dependencies between \overline{Z}_1 and the other (positive) factors in the corresponding factor frame. In diagram (b) \overline{Z}_1 does not comprise Z_3 , Z_4 , and Z_5 , rather, it contains a self-contained subset of events. In (b) \overline{Z}_1 is no longer conditionally dependent on any other factor in the respective frame and, therefore, is arbitrarily integratable into minimally sufficient conditions that contain Z_3 , Z_4 , or Z_5 . Applied to the traffic lights example, this account would yield that there are such things as absences of traffic lights in space and time that cause car crashes.

However, as shown in chapter 2, the notion of a negative event is highly problematic. Particulars in time and space are not amenable to negation – sentences only can be negated. Yet, sentences state facts, not events. While ordinary (positive) events can straightforwardly be said to involve objects that exemplify properties at times, nothing comparable is available as regards the analysis of a negative event. There are no such things as negative properties and it is far from clear in what sense objects should be involved in negative events. Absences of traffic lights are no objects exemplifying a property. On the contrary, the absence of signalling by traffic lights is there *not* being an object exemplifying the property of regulating the traffic. All in all, negative events cannot meaningfully be claimed to subsist in time and space. Such a subsistence, however, has been introduced as a necessary condition an entity has to satisfy in order qualify as an event.

These considerations demonstrate anew, why the arguments from negation⁵⁴ can be seen as the main arguments in favor of fact causation. First, the linguistic representatives of facts – sentences – are straightforwardly negatable and, second, facts are not localized in time and space. Nonetheless, as in the case of the modification of (II) discussed above, we shall only shift towards fact causation, if there is no other way to render negative factors amenable to a causal interpretation within the event causation framework.

⁵⁴ Cf. chapter 2, p. 44.

It might, of course, be argued that the instances of negative factors are not negative events, but some special kind of positive events, such that not all positive events that do not instantiate Z_1 should be seen as instances of $\overline{Z_1}$, rather, only a subclass of the non-instances of Z_1 . Obviously, this subclass has to be thus defined that not all instances of other positive factors that we want to assign to the same complex causes as $\overline{Z_1}$ are contained within that subclass. Yet, an analysis of the instances of a negative factor $\overline{Z_1}$ that stipulates them to be those events that do not instantiate factors together with which $\overline{Z_1}$ constitutes a complex cause of any effect would render a causal interpretation of negative factors outright circular. However, what other criteria are available to identify the events that instantiate a negative factor? What events that do not instantiate B in (3.47) shall be taken to instantiate \overline{B} ? I cannot think of any non-causal criterion that would accept, say, a bug crossing the intersection as an instance of \overline{B} while rejecting the instances of A . Moreover, any confinement of the event set constituting a negative factor $\overline{Z_1}$ would cast serious doubts on the first-order formalizability of sentences about factors. For any restriction of (IV) would give the negation sign an interpretation that significantly deviates from classical interpretations. Overall, therefore, the theoretical price of a modification of (III) or (IV) is far too high.

What about assumption (V)? Are conditionally dependent factors as B or \overline{B} in (V) generally redundant within minimally sufficient conditions? Indubitably, that is the case for minimally sufficient conditions as discussed in section 3.3 and as implemented in e.g. (3.52):

$$\forall x_1 \forall x_2 ((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \quad (3.52)$$

If (3.52) really expresses the sort of sufficiency that is required for a causal interpretation of a negative factor, (3.51) proves that the antecedent of (3.52) is not minimal and that \overline{B} drops out of a corresponding minimal theory of D . (3.52) essentially says that whenever there is an instance e_1 of A and an instance e_2 of \overline{B} , such that e_1 and e_2 are coincident, there is an instance e_3 of D in the same spatiotemporal frame. However, due to the vast amount of instances of \overline{B} , there are countless events of type \overline{B} that are coincident with events of type A without there ever occurring a car crash – take, for instance, people standing at the roadside while cars cross or the breathing of the drivers or take the instances of A which all instantiate \overline{B} . (3.52) is simply false relative to the factor definitions in (3.47), for the antecedent of (3.52) may well be instantiated while no cars crash. This bears an interesting consequence: (3.48) not only cannot be a minimal theory underlying the behavior of D because one of the conjunctions in the antecedent of its first conjunct is not a *minimally* sufficient condition, moreover, (3.48) does not constitute a minimal theory because that same conjunction is not even *sufficient* for D . Hence, there is something fundamentally wrong with (3.48) that goes beyond it containing conjunctions that are not minimal.

Somebody claiming that $A\overline{B}$ is a complete complex cause of D neither maintains (3.48) nor (3.52). In order to determine what the problem is with (3.48) and

(3.52), it will be useful to ascertain what is really claimed about the behavior of A , \overline{B} , and D by holding $A\overline{B}$ to be a complete complex cause of D . Cars simultaneously passing an intersection v do not crash if there merely is *at least one* instance of \overline{B} at v at the time of the passing, but only if *all* events occurring within the corresponding spatiotemporal frame are of type \overline{B} . As soon as there is just one event at v that consists in traffic lights regulating the traffic, there will be no crash. Hence, if it is held that $A\overline{B}$ is a complete complex cause of D , what is meant is that whenever there is an instance of A while *there is no* coincident instance of B , an event of type D occurs. This is not expressed by (3.52), but by (3.53) or equivalently by (3.54):⁵⁵

$$\forall x_1((Ax_1 \wedge \neg \exists x_2(Bx_2 \wedge Kx_1x_2)) \rightarrow \exists y(Dy \wedge x_1 \neq y \wedge Rx_1y)) \quad (3.53)$$

$$\forall x_1((Ax_1 \wedge \forall x_2 \neg (Bx_2 \wedge Kx_1x_2)) \rightarrow \exists y(Dy \wedge x_1 \neq y \wedge Rx_1y)). \quad (3.54)$$

(3.53) and (3.54) deviate from (3.52) in twofold ways. First, whereas the antecedent of (3.52) is satisfied by any coincident instantiation of A and \overline{B} , the antecedent of (3.53) is only satisfied if all events that are coincident with an instance of A are of type \overline{B} . Second, whereas (3.52) requires the instances of \overline{B} to be related in terms of R to the car crash, (3.53) and (3.54) do not impose any constraints as regards the relation between the (absent) instances of B and the crash. While the first deviation should be sufficiently accounted for by the considerations above, the second difference between (3.53) and (3.54) on the one hand and (3.52) on the other stems from the fact that nothing with respect to the spatiotemporal relationship between *inexistent* – and thus *not* spatiotemporally located – events of type B and car crashes occurring in time and space can reasonably be stipulated.

As (3.53) and (3.54), in contrast to (3.52), accurately express what is implied by $A\overline{B}$ being a complete complex cause of D , some modifications of the syntax of minimal theories are called for. More specifically, the mentioned twofold deviations from (3.52) induce the following two modifications: First, quantifiers may appear within conjunctions stating minimally sufficient conditions and, second, the introduction of additional variables is called for, which exclusively appear in the antecedent of a minimal theory. Incorporating these modifications yields the following minimal theory of D , for which, in light of (3.53) and (3.54) two equivalent expressions are available:

$$\begin{aligned} &\forall x_1 \forall x_2 ((Ax_1 \wedge \neg \exists x_3 (Bx_3 \wedge Kx_1x_3)) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \wedge \\ &\forall y (Dy \rightarrow \exists x_1 \exists x_2 (((Ax_1 \wedge \neg \exists x_3 (Bx_3 \wedge Kx_1x_3)) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \end{aligned} \quad (3.55)$$

$$\begin{aligned} &\forall x_1 \forall x_2 ((Ax_1 \wedge \forall x_3 \neg (Bx_3 \wedge Kx_1x_3)) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \wedge \\ &\forall y (Dy \rightarrow \exists x_1 \exists x_2 (((Ax_1 \wedge \forall x_3 \neg (Bx_3 \wedge Kx_1x_3)) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge Rx_1x_2y)) \end{aligned} \quad (3.56)$$

⁵⁵ The equivalence of (3.53) and (3.54) is guaranteed by the definition of “ $\neg \exists$ ”.

Since (3.56) with its universally quantified antecedent of the first conjunct is syntactically closer to the form of minimal theories as encountered so far, we shall in the following focus on (3.56). Clearly however, this choice is arbitrary. Nonetheless, (3.56) exhibits the deviation from (3.48) in a more transparent way. What really is negated when causal relevance for crashes is attributed to the absence of a signalling of traffic lights is not the factor B itself, but B being instantiated coincidentally with cars passing the respective intersection. For brevity, we shall nevertheless continue to say that minimal theories as (3.56) attribute causal relevance to negative factors.

Of course, requiring minimal theories that attribute causal relevance to negative factors to be of the form of (3.56) only makes sense if (3.56) is free from the defects of (3.48). Clearly, cars simultaneously passing intersection v while people are standing at v , contrary to (3.48), does not satisfy the antecedent of (3.56) or of (3.54). In order to satisfy (3.54), no event of type B is allowed to occur coincidentally with two cars passing simultaneously at v . Thus, (3.56) and (3.54) are considerably stronger than (3.48) and (3.52). Other than (3.48), the antecedent of (3.56) in fact contains two sufficient conditions of D . However, are these sufficient conditions both minimal as well? Does the antecedent of (3.53) and (3.54) mention a minimally sufficient condition of D , even in light of the fact that \overline{B} conditionally depends on A ? Even relative to a reflexive K , this question can be answered in the affirmative. While every instance of A is an instance of \overline{B} , from A being instantiated it does not follow that there are no events of type B that occur coincidentally with a respective instance of A . The following implication does *not* hold:

$$\forall x(Ax \rightarrow \neg Bx) \rightarrow \forall x_1(Ax_1 \rightarrow \forall x_2 \neg(Bx_2 \wedge Kx_1x_2)).$$

In contrast, (3.57) holds:⁵⁶

$$\begin{aligned} (\forall x_1(Ax_1 \wedge \forall x_2 \neg(Bx_2 \wedge Kx_1x_2)) \rightarrow \exists y(Dy \wedge x_1 \neq y \wedge Rx_1y)) \wedge \forall x Kxx \wedge \\ \forall x Rxx \wedge \forall x(Ax \rightarrow \neg Bx) \not\rightarrow \forall x_1(Ax_1 \rightarrow \exists y(Dy \wedge x_1 \neq y \wedge Rx_1y)). \end{aligned} \quad (3.57)$$

That means, even relative to a reflexive interpretation of K , $\forall x_2 \neg(Bx_2 \wedge Kx_1x_2)$ does not become redundant if it is part of a sufficient condition that contains A .

What has been said thus far with respect to the traffic lights example can be generalized to all minimal theories that attribute causal relevance to absences and negative factors. Minimal theories that mirror complex causes containing negative factors are of the logical form of (3.56). Strictly speaking, they do not contain negative factors per se, rather, expressions of type (3.56) negate the corresponding positive factor being coincidentally instantiated with other factors in a respective minimally sufficient condition. For reasons of simplicity and transparency, we shall nonetheless stick to our abbreviated notation. Therefore,

$$\overline{AB} \vee AC \Rightarrow D$$

⁵⁶ A proof is provided in the appendix (p. 265).

is to be seen as an abbreviation of (3.56), not of (3.48).

If a complex cause contains more than one positive factor along with at least one negative factor, further adjustments are called for. In order to illustrate this, let us slightly modify our example. Suppose, over and above an instance of A and the absence of B what is required for a car crash to occur is the street being wet.

$$E = \{x : x \text{ is a wet street}\}$$

Hence, assume that $A\overline{B}E$ is a complete complex cause of D . What does this mean as regards the behavior of these four factors? First, crashes occur when it is the street on which the cars pass that is wet. Thus, if A and E are not coincidentally instantiated, D is not determined to take place. Furthermore, the traffic on the respective intersection is not to be regulated by traffic lights. That is, crashes occur if all events that are coincident with instances of AE are of type \overline{B} . That means, the positive factors are required to be coincidentally instantiated and the negative factor is required to be coincidentally instantiated with the positive factors. This gives rise to two different coincidence relations, one with two arguments and one with three. By resorting to two such relations, $K_1x_1x_2$ and $K_2x_1x_2x_3$, the dependency among $A\overline{B}E$ and D can be expressed as follows:

$$\begin{aligned} \forall x_1 \forall x_2 ((Ax_1 \wedge Ex_2 \wedge K_1x_1x_2 \wedge \forall x_3 \neg (Bx_3 \wedge K_2x_1x_2x_3)) \rightarrow \\ \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge R_{x_1x_2y})) \end{aligned} \quad (3.58)$$

Both K_1 and K_2 can be treated as ordinary non-redundant parts of sufficient conditions and, accordingly, do not need to be given fixed interpretations (cf. p. 97). This finding can be generalized. If causal relevance is attributed to negative factors contained in a complex cause involving at least two positive factors, a second coincidence relation is called for which holds among each negative factor and the positive factors in a corresponding complex cause. If K_1 is n -ary, K_2 is $n + 1$ -ary. In light of (3.58), our abbreviated notation has to be further accommodated. For the purpose of not having to explicitly mention that instances of complex causes that contain negative factors are required to be instantiated in terms of corresponding coincidence relations K_1 and K_2 , we adopt the following convention.

$$\begin{aligned} Z_1 \dots Z_i \overline{Z_{i+1}} \dots \overline{Z_n} =_{df} Z_1x_1 \wedge \dots \wedge Z_ix_i \wedge K_1x_1 \dots x_i \wedge \\ \forall x_{i+1} \neg (Z_{i+1}x_{i+1} \wedge K_2x_1 \dots x_ix_{i+1}) \wedge \dots \wedge \forall x_n \neg (Z_nx_n \wedge K_2x_1 \dots x_ix_n) \end{aligned} \quad (3.59)$$

As in the case of concatenations of merely positive factors the quantifiers that bind the variables on the right-hand side of (3.59) can be left unspecified, because we will only use this abbreviated notation in connection with “ \mapsto ” and “ \Rightarrow ”, both of which uniquely determine the sort of quantifiers involved. Therefore, the context in which expressions of type $Z_1 \dots Z_i \overline{Z_{i+1}} \dots \overline{Z_n}$ appear will always clarify the nature of the quantifiers binding the free variables in (3.59). Given this notational

convention, the minimal theory for our extended traffic lights example can be stated thus:

$$AE\bar{B} \vee AC \Rightarrow D. \quad (3.60)$$

All of these considerations show that it is assumption (V) of the argument against the integration of negative factors in minimal theories that has to be accommodated. It is not generally the case that one of two conditionally dependent factors is redundant in sufficient conditions that contain the other factor. (V) only holds for positive factors. In order to account for our revised treatment of negative factors, (V) needs to be modified in the following way:

(V') If $\forall x(\Gamma(x) \rightarrow \Sigma(x)) \wedge \neg\forall x(\Sigma(x) \rightarrow \Gamma(x))$, $\Sigma(x)$ is redundant in a minimally sufficient condition containing $\Gamma(x)$,

where $\Gamma(x)$ and $\Sigma(x)$ represent any two formulas containing at least one free occurrence of x . Regardless of (3.49), \bar{B} does not become redundant in a minimal theory of D , for (3.61) does not hold for the example at hand:

$$\forall x_1(Ax_1 \rightarrow \forall x_2\neg(Bx_2 \wedge Kx_1x_2)). \quad (3.61)$$

Thus, the first conjunct of (V') is violated by the traffic lights example. All the other assumptions involved in the argument discussed in this section can be sustained without modifications.

3.6.5 Simple vs. Complex Minimal Theories (I)

Sections 3.3 and 3.4 have shown that conditionally dependent factors cannot be part of the same minimal theory, and from section 3.6.3 we know that factors on different levels of specification cannot be part of the same minimal theory. This section will first be concerned with the question as to whether *causally dependent* factors can be contained in the same minimal theory, and subsequently introduce complex minimal theories that represent complex causal structures.

Consider the causal structure depicted by graph (a) in figure 3.4. A and B are part of two different minimally sufficient conditions of C and those conditions are

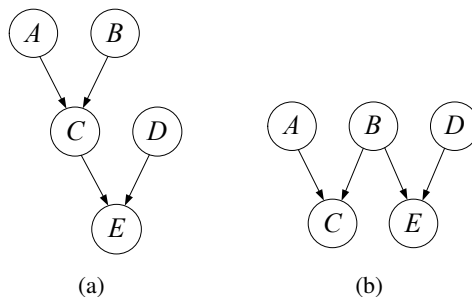


Fig. 3.4: Two complex causal structures that have to be represented by complex minimal theories.

contained in a minimally necessary condition of C . The same applies to C and D with respect to E . This yields two minimal theories, one for C and one for E :

$$AX_1 \vee BX_2 \vee Y_C \Rightarrow C \quad (3.62)$$

$$CX_3 \vee DX_4 \vee Y_E \Rightarrow E. \quad (3.63)$$

Assuming that A , B , C , and D remain part of (3.62) and (3.63) across all factor frame extensions renders those minimal theories amenable to a causal interpretation. Moreover, A and B are each part of a minimally sufficient condition of E . Relative to a respective spatiotemporal frame: Whenever AX_1 or BX_2 are instantiated, there is an event of type C ; if, furthermore, the factors in X_3 are instantiated, there also is an event of type E . That means, AX_1X_3 and BX_2X_3 are each minimally sufficient for E . This finding is generalizable. Causes are not only part of minimally sufficient conditions of their *direct*, but also of their *indirect* effects. This is essentially guaranteed by the transitivity of R .⁵⁷ For, if instances of AX_1 or BX_2 are related in terms of R to an instance of C and this event, in turn, is related in terms of R to an instance of E , it follows that the instances of AX_1 and BX_2 are properly spatiotemporally related to the corresponding event of type E .

In order to determine whether minimal theories can contain causally dependent factors, we need to examine whether a minimally sufficient condition comprising C can be extended by A or B and whether a minimally necessary condition involving C can be extended by AX_1X_3 or BX_2X_3 . Clearly, neither ACX_3 nor BCX_3 are minimally sufficient for E , for there is a proper part of both ACX_3 and BCX_3 that is sufficient for E : CX_3 . Are (3.64) or (3.65) minimally necessary conditions of E ?

$$CX_3 \vee AX_1X_3 \vee DX_4 \vee Y_E \quad (3.64)$$

$$CX_3 \vee BX_2X_3 \vee DX_4 \vee Y_E \quad (3.65)$$

The answer must be in the negative, for both (3.64) and (3.65) have a necessary proper part: the antecedent of (3.63).⁵⁸ Hence, the antecedents of minimal theories do not contain causally dependent factors. Minimal theories, as defined thus far, cannot represent causal chains as depicted by graph (a) of figure 3.4.

Furthermore, section 3.5 has established that, even though factor C of graph (b) in figure 3.4 might be integrated within a minimal theory of E , given an appropriate factor frame, C will be dropped from that minimal theory upon extensions of the respective factor frame. The syntax of minimal theories has been devised such that the consequents of minimal theories contain one factor only. Both the non-persistent integratability of C in a minimal theory of E , i.e. the fact that causally

⁵⁷ Cf. section 3.2 above. Of course, this consequence might be blocked if R were taken to be a non-transitive relation. In chapter 4 a detailed discussion of the merits and defects of a non-transitive R will show that a transitive R is ultimately preferable.

⁵⁸ If AX_1X_3 and BX_2X_3 are disjunctively integrated into the same necessary condition of E along with CX_3 , the resulting necessary condition contains two necessary proper parts, which renders the minimalization of that necessary condition ambiguous. This problem will be addressed in chapter 4.

interpretable minimal theories do not contain epiphenomenally dependent factors, and the syntactical constraints of minimal theories demonstrate that causally interpretable minimal theories, as defined thus far, cannot represent epiphenomenal structures as the one depicted in graph (b).

A representation of complex causal structures calls for the composition of complex minimal theories from simple minimal theories, the latter being theories as in (3.62) and (3.63). In order to illustrate how simple minimal theories can be put together to complex ones, consider anew graph (a) of figure 3.4. Relative to this causal structure, A and B are and remain part of a minimal theory of C and C and D are and remain part of a minimal theory of E . In our abbreviated notation this is straightforwardly expressed by a mere conjunction of (3.62) and (3.63):

$$(AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (CX_3 \vee DX_4 \vee Y_E \Rightarrow E). \quad (3.66)$$

In an analogous way the epiphenomenon depicted in graph (b) of figure 3.4 can be represented by means of a conjunction of two simple minimal theories. Relative to this epiphenomenal structure, A and B are and remain part of a minimal theory of C and B and D are and remain part of a minimal theory of E . This is expressed by the following complex minimal theory:

$$(AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (BX_3 \vee DX_4 \vee Y_E \Rightarrow E). \quad (3.67)$$

All in all, complex minimal theories result from conjunctively combining simple minimal theories. However, not every conjunction of simple minimal theories constitutes a complex theory. Contrary to (3.66) and (3.67), (3.68) and (3.69) are no complex minimal theories.

$$(AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (GX_3 \vee HX_4 \vee Y_E \Rightarrow E) \quad (3.68)$$

$$(AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \quad (3.69)$$

Complex minimal theories specify coherent complex causal structures. (3.68) does not do that. It represents *two* disconnected causal contexts. For a conjunction of two simple minimal theories Φ and Ψ to exhibit a coherent structure at least one factor must be contained both in Φ and in Ψ . Moreover, complex minimal theories are minimal in the sense that they do not contain redundant conjuncts. (3.69) does not satisfy this minimality condition as one of the conjuncts is redundant. Complex minimal theories represent causal chains or epiphenomena. Simple theories with identical consequents cannot constitute a complex minimal theory if conjunctively combined, for (3.69) neither specifies a causal chain nor an epiphenomenon.

The notion of a complex minimal theory shall be defined inductively by first explicitly introducing the notion of a simple minimal theory:

Simple and complex minimal theories (I):

1. A double-conditional with an antecedent consisting of a minimally necessary disjunction of minimally sufficient conditions and a consequent consisting of a single factor is a *simple* minimal theory of that consequent.

2. Every simple minimal theory is a minimal theory.
3. A conjunction of two minimal theories Φ and Ψ is a minimal theory iff
 - (a) at least one factor in Φ is part of Ψ and
 - (b) Φ and Ψ do not have an identical consequent.
4. Minimal theories that are not simple are *complex*.

This definition is to be seen as first proposal. It still suffers from an important defect that will be the central subject of the next chapter.

3.6.6 Causal Relevance Defined

Notwithstanding the fact that the notion of a complex minimal theory will have to be further adapted in the following chapter, we now have assembled all theoretical means necessary to bring our analysis of causal relevance to its final form.

As in the case of simple minimal theories, not all factors in complex minimal theories are causally interpretable. A factor that is part of a complex minimal theory is only causally interpretable if it stays part of that minimal theory across all extensions of the corresponding factor frame. The introduction of the notions of simple and complex minimal theories induces adjustments of MT_d^* and MT_i^* . In particular, indirect causal relevance can now be accounted for independently of MT_d^* . Moreover, section 3.6.4 has established that MT^* spells out causal relevance for *positive* factors only, which shall be made explicit in the following – along with a suitable modification of MT_n^* .

Direct causal relevance (MT_d): A factor A (\bar{A}) is directly causally relevant for a positive factor B iff

- (a) A (\bar{A}) is part of a *simple* minimal theory Φ of B
- (b) A (\bar{A}) stays part of Φ across all extensions of its factor frame.

Indirect causal relevance (MT_i):⁵⁹ A factor A (\bar{A}) is indirectly causally relevant for a positive factor B iff there is a sequence S of factors Z_1, Z_2, \dots, Z_n , $n \geq 3$, such that

- (a) $A = Z_1$ ($\bar{A} = \bar{Z}_1$) and $B = Z_n$
- (b) for each i , $1 \leq i < n$: Z_i is part of a simple minimal theory of Z_{i+1}
- (c) the conjunction of the minimal theories of Z_2, Z_3, \dots, Z_n constitutes a complex minimal theory Ψ
- (d) the factors in Ψ stay part of Ψ across all extensions of Ψ 's factor frame.

Causal relevance (MT): A factor A (\bar{A}) is causally relevant for a positive factor B iff A (\bar{A}) is directly causally relevant for B in terms of MT_d or indirectly causally relevant for B in terms of MT_i .

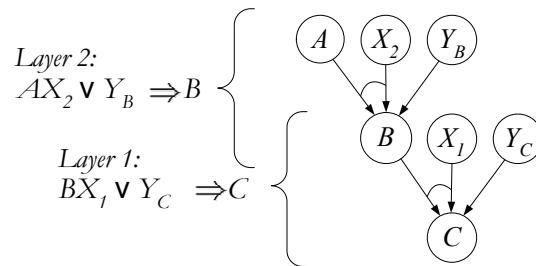


Fig. 3.5: A complex causal graph consisting of two layers. Minimal theories correspond to single layers of causal graphs.

Causal relevance for negative factors (MT_n): A factor A (\bar{A}) is causally relevant for a negative factor \bar{B} iff the negation of A (\bar{A}) is causally relevant for B according to MT.

Indirect causal relevance can alternatively – and maybe more illustratively – be analyzed by graphical means. The antecedents of simple minimal theories only contain factors that are – relative to a respective factor frame – directly causally relevant to their consequents. The transitivity of causal relevance (TR), however, allows for deriving indirect from direct causal relevance. This is plainly illustrated by the graph notation. Simple minimal theories describe single *layers* of complex causal graphs. By a *layer* of a causal graph G we shall refer to a subgraph⁶⁰ G' of G , i.e. $G' \subseteq G$, containing a maximal number of vertices of G such that G' consists of two sets M and N of vertices, the vertices in M and N are not mutually adjacent, there exists at least one path from every vertex in M to at least one vertex in N , and M merely contains heads of edges and N merely tails.⁶¹

Causal graphs consisting of two layers or more will be referred to as *chain-graphs*. To each layer of a chain-graph – as the one in figure 3.5 – there corresponds a simple minimal theory. Layers of causal graphs are numbered starting with the lowermost layer, i.e. with the layer whose tail does not dominate any other factors in the graph. In this sense, the second layer of the graph in figure 3.5 is constituted by the factor sets $\{A, X_2, Y_B\}$ and $\{B\}$, while the first layer consists of the sets $\{B, X_1, Y_C\}$ and $\{C\}$. Each of the two layer constituting factor sets corresponds to antecedent and consequent of the minimal theories assigned to these layers in figure 3.5.

By means of causally interpretable minimal theories chain-graphs can be built up layer by layer. While direct relevance only can be read off a simple minimal theory, a chain-graph G , such that a causally interpretable minimal theory corresponds to every layer in G , straightforwardly reveals indirect relevancies. A factor

⁵⁹ The notion of a complex minimal theory resorted to in MT_i is to be understood in terms of section 4.5.2 below.

⁶⁰ For details on the notion of a subgraph see Bang-Jensen and Gutin (2001), ch. 1.

⁶¹ Cf. section 2.3.2.

A in G is indirectly causally relevant for a factor C iff there is a *trail* in G such that A is the head of that trail and C its tail – a trail in a graph G being a sequence S of vertices Z_1, Z_2, \dots, Z_n , $n \geq 3$, in G such that between every pair of adjacent vertices of S there exists a path and any Z_i is the tail of the path whose head is Z_{i-1} . That means, between all vertices Z_i and Z_{i+1} in S there exists a directed edge from Z_i to Z_{i+1} . The head of a trail is the first vertex in S , its tail the last vertex in S .

In the graph of figure 3.5 there exists a trail from A to C , namely $A \longrightarrow B \longrightarrow C$.⁶² This allows for an inference to the indirect causal relevance of A for C . Hence, alternatively to MT_i indirect causal relevance is definable by graphical means:

Indirect causal relevance and chain-graphs: A factor A is indirectly causally relevant for a factor C iff there exists a chain-graph G such that a causally interpretable simple minimal theory corresponds to every layer of G and there exists a trail from A to C in G .

3.7 Causation Among Events – Singular Causation

Thus far we have mainly been concerned with type level (or general) causation, which, as shown in the introductory remarks to this chapter, is the primary analysandum of a regularity account of causation. A regularity theory professes that causal dependencies among singular events, i.e. token level or singular causation, are not accessible without recourse to regularities on type level. Causation among events is to be derived from causal relevance among factors. This section addresses the question as to how singular causation follows from causal relevance, i.e. it proposes a definition of singular causation based on MT .

Prima facie, defining singular causation on the basis of MT seems to be utterly straightforward:

Singular causation (SC):* An event a is a cause of an event b iff a instantiates a factor A or its negation \bar{A} and b instantiates a factor B , such that

- (a) A (\bar{A}) is part of a minimal theory Φ of B ,
- (b) Φ is causally interpretable according to MT ,
- (c) $a \neq b$ and a and b occur within the same spatiotemporal frame, and
- (d) a is coincident with other events that instantiate a minimally sufficient condition X of B which is part of Φ and contains A (\bar{A}).

SC^* is defective in one important respect. Yet, before this defect is discussed, let us illustrate the significant qualities of this straightforward regularity theoretic

⁶² Long right arrows as “ \longrightarrow ” shall symbolize arrows in graphs.

analysis of singular causation. Suppose (3.70) is a causally interpretable complete minimal theory of B .

$$\begin{aligned} & \forall x_1 \forall x_2 ((Ax_1 \wedge Dx_2 \wedge Kx_1x_2) \vee (Cx_1 \wedge Fx_2 \wedge Kx_1x_2) \rightarrow \exists y (By \wedge x_1 \neq y \wedge x_2 \neq y \wedge R_{x_1x_2}y)) \wedge \\ & \forall y (By \rightarrow \exists x_1 \exists x_2 (((Ax_1 \wedge Dx_2 \wedge Kx_1x_2) \vee (Cx_1 \wedge Fx_2 \wedge Kx_1x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge R_{x_1x_2}y)) \end{aligned} \quad (3.70)$$

Relative to SC^* an event a is a cause of an event b iff there is a model of (3.70) in which AD is instantiated by two coincident events, such that the instance of A is a , while b instantiates B , and moreover $a \neq b$ and Rab . Thus, for instance:⁶³

$$\begin{aligned} I &= \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\} \\ \mathfrak{S}(a) &= c_1 \\ \mathfrak{S}(b) &= c_2 \\ \mathfrak{S}(A) &= \{c_1, c_5\} \\ \mathfrak{S}(B) &= \{c_2, c_4\} \\ \mathfrak{S}(C) &= \{c_3, c_6\} \\ \mathfrak{S}(D) &= \{c_7, c_8\} \\ \mathfrak{S}(F) &= \{c_9, c_{10}\} \\ \mathfrak{S}(K) &= \{(c_1, c_7), (c_7, c_1), (c_5, c_8), (c_8, c_5)\} \\ \mathfrak{S}(R) &= \{\dots, (c_1, c_7, c_2), (c_5, c_8, c_4), \dots\} \end{aligned} \quad (\mathfrak{S}_1)$$

In many cases, this clear-cut analysis of what it means for a to be a cause of b perfectly mirrors common causal judgements. Suppose there are exactly two complex causes of gastrospasm (B): gastric ulcer (A) in combination with smoking (D) and gastritis (C) in combination with excessive caffeine consumption (F). Hence, (3.70) shall be taken to represent the complete causal structure behind suffering from gastrospasm. Now, we interpret the events c_1 to c_4 in the domain of (\mathfrak{S}_1) as follows:

$$\begin{aligned} c_1 &= \text{Hogan's gastric ulcer on October 5, 2004} \\ c_2 &= \text{Hogan's gastrospasm on October 5, 2004} \\ c_3 &= \text{Fennella's gastritis on October 5, 2004} \\ c_4 &= \text{Fennella's gastrospasm on October 5, 2004} \end{aligned}$$

(3.70) identifies Hogan's gastric ulcer as a cause of Hogan's gastrospasm in a forthright way. c_1 is a cause of c_2 because Hogan does not only suffer from gastric ulcer, but in addition is assigned to be a smoker by (\mathfrak{S}_1) – c_1 is coincident with c_7 , which is of type D . Relative to (\mathfrak{S}_1) , Fennella's gastritis, however, is not determined to be a cause of Fennella's gastrospasm, for in combination with c_3 there is

⁶³ I specifies the domain of quantification and \mathfrak{S} assigns entities of the domain to singular terms and predicates. For details on interpretations of first-order formulas see e.g. Ebbinghaus, Flum, and Thomas (1998), §§1-3. “...” in $\mathfrak{S}(R)$ stands for additional triples which, taken together, meet the requirements imposed on interpretations of R (cf. page 93 above).

no coincident instance of F . Rather, Fennella's gastrospasm was caused by events c_5 and c_8 , which, presumably, were Fennella's gastric ulcer and her smoking. For an event a to cause an event b it is thus not sufficient that a instantiates any factor in a minimal theory of B . Rather, a needs to instantiate a factor in a minimally sufficient condition, whose factors are *all* instantiated coincidentally with a .

The *prima facie* analysis of singular causation (SC^*) has further important qualities: It mirrors the relational properties attributed to singular causation in section 2.3.1. In order to see this, consider the following interpretation of (3.70):⁶⁴

$$\begin{aligned}
 I &= \{c_1, c_2, c_3, c_4, c_5\} \\
 \mathfrak{S}(A) &= \{c_1, c_2\} \\
 \mathfrak{S}(B) &= \{c_1, c_2, c_5\} \\
 \mathfrak{S}(C) &= \{c_2, c_3\} \\
 \mathfrak{S}(D) &= \{c_3, c_4\} \\
 \mathfrak{S}(F) &= \{c_1, c_5\} \\
 \mathfrak{S}(K) &= \{(c_1, c_1), (c_2, c_2), (c_3, c_3), (c_4, c_4), (c_5, c_5) \\
 &\quad (c_1, c_3), (c_3, c_1), (c_2, c_4), (c_4, c_2)\} \\
 \mathfrak{S}(R) &= \{\dots, (c_1, c_3, c_2), (c_2, c_4, c_1), (c_3, c_1, c_5), \dots\}
 \end{aligned} \tag{\mathfrak{S}_2}$$

(\mathfrak{S}_2) is a model of (3.70): For every instance of $AD \vee CF$ there is a different instance of B in the same spatiotemporal frame and for every instance of B there is a different instance of $AD \vee CF$ in the same spatiotemporal frame. This model of (3.70) has the following interesting peculiarity: For event c_1 which is of type A and coincident with c_3 which is of type D there is a properly located instance of B , namely c_2 . Event c_2 is also an instance of A and, moreover, coincident with c_4 which is of type D . For this coincident occurrence of c_2 and c_4 there again is a properly located instance of B , namely c_1 . Accordingly, SC^* identifies c_1 as a token level cause of c_2 and c_2 as a token level cause of c_1 . That means, on behalf of SC^* (\mathfrak{S}_2) describes a token level feedback. In contrast, interpretation (\mathfrak{S}_1) does not feature such a causal feedback among event. Thus, SC^* is compatible with token level feedbacks, but does not require such feedbacks to obtain, i.e. it determines singular causation to be *non-symmetric*. Furthermore, even though c_1 causes c_2 which, in turn, causes c_1 relative to (\mathfrak{S}_2), neither c_1 nor c_2 are determined to cause themselves by SC^* . In order for an event a to be the cause of an event b according to SC^* , a and b must be different events. That means, SC^* yields an *irreflexive* notion of singular causation. Finally, (\mathfrak{S}_2) illustrates that SC^* determines singular causation to be non-transitive. As mentioned above, relative to (\mathfrak{S}_2) c_1 is a cause of c_2 and c_2 is a cause of c_1 , but c_1 is not identified to cause itself by SC^* . Hence, singular causation is not transitive. In contrast, SC^* not only determines c_2 to be a cause of c_1 in (\mathfrak{S}_2) and c_1 to be a cause of c_5 , but moreover stipulates that c_2 is a

⁶⁴ Note that factors A to F here are not assumed to be interpreted in terms of the gastrospasm example. Again, “...” in $\mathfrak{S}(R)$ stands for additional triples which, taken together, meet the requirements imposed on interpretations of R (cf. page 93 above).

cause of c_5 . Accordingly, singular causation is not intransitive subject to analysis SC^* . Therefore, SC^* provides a *non-transitive* notion of singular causation.

Despite all these qualities, this seemingly obvious analysis of singular causation has a serious drawback. SC^* fails in case of causally relevant *negative* factors. Consider anew the example of a car crash that is due to the absence of traffic lights at an intersection.⁶⁵ There are exactly two causes for car crashes at intersections (D): Missing traffic lights (\overline{B}) in combination with two cars heading towards each other (A) and drunken driving across the intersection (C) in combination with A . As we have seen in the previous section, (3.56) is the minimal theory representing this causal structure.

$$\begin{aligned} & \forall x_1 \forall x_2 ((Ax_1 \wedge \forall x_3 \neg (Bx_3 \wedge Kx_1 x_3)) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1 x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \wedge \\ & \forall y (Dy \rightarrow \exists x_1 \exists x_2 (((Ax_1 \wedge \forall x_3 \neg (Bx_3 \wedge Kx_1 x_3)) \vee (Ax_1 \wedge Cx_2 \wedge Kx_1 x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \end{aligned} \quad (3.56)$$

Let us assume that events are related in terms of R to events of type D if they are located within 20 meters of the intersection at the time of an accident. Now, we look at a concrete car crash d , say, Hogan's and Fennella's unfortunate crashing on October 6, 2004, which is preceded by Hogan and Fennella driving towards each other – event a . At the time of d Shamus is standing at the intersection of the accident waiting for Fennella to drive by so that he can cross the street. To Shamus' standing at the intersection we shall refer by s . Event s apparently instantiates \overline{B} – Shamus' waiting is not a traffic light signalling. Furthermore, assume s is coincident with a . Hence, s satisfies all the conditions imposed on singular causation by SC^* : It instantiates a factor in a minimal theory of D such that it is coincident with events that instantiate the other factors of a minimally sufficient condition of D and occurs in the same spatiotemporal frame as d . Relative to this analysis of singular causation, Shamus' standing at the intersection would thus have to be identified as a cause of Hogan's and Fennella's crash. This clearly is not acceptable. People are not held responsible for crashes if they incidentally happen to stand at the site of the crash. Of course, s is by far not the only instance of \overline{B} that is coincident with a and related in terms of R to d . There are bugs crawling across the respective intersection, birds flying above the crash site, or molecules twirling around in the pavement at the intersection etc. All of these events instantiate \overline{B} and thus would have to be causally interpreted according to SC^* .

Positive and negative factors differ fundamentally with respect to the causal interpretability of their instances. In sections 2.2.1 and 3.6.4 we have determined singular causation to be a relation that subsists between events, i.e. entities in time and space. While such entities are readily available in case of positive factors, there is no such thing as a negative event. Moreover, the discussion in the previous section has shown that what is negated when causal relevance is attributed to absences is not a factor per se, but a factor being instantiated coincidentally with a given number of other factors. Hence, the antecedent of a minimal theory containing a negative

⁶⁵ Cf. sections 2.2.1 and 3.6.4.

factor in cause position as (3.56) is not satisfied by merely *one* instance of \overline{B} being coincident with an instance of A . Rather, what it takes for the antecedent of (3.56) to be satisfied is that *all* events that are coincident with events of type A are instances of \overline{B} . In this regard, (3.56) significantly diverges from (3.70). While every coincident instantiation of a disjunct in a minimal theory as (3.70) – e.g. Hogan’s gastric ulcer and smoking on October 5, 2004 – satisfies the antecedent of such a theory, it takes much more for the antecedent of theories of type (3.56) to be satisfied. Shamus’ standing at the intersection at the time of Hogan’s and Fennella’s simultaneous passing does not satisfy the antecedent of (3.56). What satisfies the antecedent of (3.56) is a simultaneous passing of cars while *there are no* coincident signalings of traffic lights. The latter is not an *event* in space and time, but can be interpreted as a *fact*. However, this does not compromise our commitment to event causation. What determines the ontology of causation is the domain of quantification, which in our case consists of events only, not of localities in time and space or objects or even facts. Causally interpreted absences do not subsist in time and space as events do. Therefore, they are not suited as relata of singular causation. This is captured by our formalization of causal relevance statements by means of minimal theories.

All in all, only positive factors have instances that are amenable to causal interpretations. Relata of singular causation are entities in time and space that happen to satisfy disjuncts of minimal theories. Expressions as $\forall x_3 \neg (Bx_3 \wedge Kx_1x_3)$ in (3.56) are not rendered true by entities in time and space, but by facts. SC^* must be amended as follows:

Singular causation (SC): An event a is a cause of an event b iff a instantiates a *positive* factor A and b instantiates a factor B , such that

- (a) A is part of a minimal theory Φ of B ,
- (b) Φ is causally interpretable according to MT,
- (c) $a \neq b$ and a and b occur within the same spatiotemporal frame, and
- (d) a is coincident with other events that instantiate a minimally sufficient condition X of B which is part of Φ and contains A .

While instances of causally relevant positive factors are amenable to causal interpretations, instances of causally relevant negative factors are not. Negative factors can be attributed causal relevance in the way discussed in the previous section, yet their instances are not relata of singular causation. Relata of singular causation have to conform to two requirements: (1) They are entities that satisfy disjuncts in the antecedents of minimal theories and (2) they are spatiotemporally located. What satisfies condition (1) in case of minimal theories as (3.56) does not satisfy condition (2).

This analysis of singular causation might be criticized for resting on an arbitrary distinction between positive and negative factors. Statements (57) and (58) on

page 120 illustrate that there are factors defined by contradictory predicates such that one seemingly positive factor as

$$H = \{x : x \text{ is a condition of being dry}\}$$

is identical to the negation \bar{J} of a factor as

$$J = \{x : x \text{ is a condition of being wet}\}.$$

Discarding the fuzziness of the threshold between wetness and dryness, an event instantiates H if and only if it instantiates \bar{J} . Therefore, H is part of a minimally sufficient condition of any effect iff \bar{J} is part (of another)⁶⁶ minimally sufficient condition of the same effect. Now take a concrete causal process as dryness (H) of a match being a causally relevant factor for the match to catch fire when struck – refer to this effect as L . Equivalently it might be said that the absence of wetness (\bar{J}) on a struck match is causally relevant for it to light. Thus we have two minimal theories of L , one containing H and one containing \bar{J} . Suppose e_1 to be a singular event instantiating H and \bar{J} , while e_2 shall be taken to be a lighting of a match that occurs in the same spatiotemporal frame as e_1 . Moreover, coincidentally with e_1 all other factors of the corresponding minimally sufficient condition of L shall be instantiated. Against this background, SC identifies e_1 as a cause of e_2 only if e_1 is seen as an instance of H . If e_1 is taken to instantiate \bar{J} , it is not determined to be a cause of e_2 . The choice as to whether H or \bar{J} are held to be instantiated by e_1 is completely arbitrary. This seems to suggest that whether or not e_1 is amenable to a causal interpretation relative to SC depends on the description chosen for e_1 .

Clearly, an event a being a cause of an event b is not a matter of how a is typecast. Rather, it is an objective matter of fact. Indeed, SC does not have any implications to the contrary. MT determines A to be causally relevant for B if *there is at least one* minimal theory of B that contains A in its antecedent such that A remains part of that theory across all factor frame extensions. Furthermore, a necessary condition for SC to identify a as a cause of b is *there being at least one* minimal theory of B that contains a positive factor – across all factor frame extensions – in its antecedent that is instantiated by a . This condition is satisfied in case of the match example above: There is a minimal theory that comprises a positive factor (H) in its antecedent which is instantiated by e_1 . The fact that there is another minimal theory that does not contain this positive factor is of no concern as to whether e_1 is a cause of e_2 or not. It might, of course, happen that a minimal theory involving H is not *known* at a certain moment of investigating into the causes of events of type L . This lack of knowledge, however, is of epistemic interest only and does not hamper a correct identification of e_1 as a cause of e_2 on conceptual grounds.

There is a second, yet very similar, problem that arises from the vagueness of the notions of a positive and of a negative factor. As (61) and (62) illustrate, it is

⁶⁶ Identical factors cannot be plurally contained in one and the same minimally sufficient condition, for one of two occurrences of identical factors will always be redundant.

often unclear whether a causal process is to be described by means of positive or negative factors.

- (61) Drowsiness is causally relevant for stumbling.
- (62) Absence of responsiveness is causally relevant for stumbling.

Depending on what factor frame is chosen to describe the causal process characterized in (61) and (62), a certain event may be identifiable as a cause of a concrete stumbling or not. Yet again, this problem is of epistemic concern only.⁶⁷ Whether an event is *identifiable* as a cause relative to a given factor frame and whether this event *actually is* a cause are two different questions – one of epistemic and the other of conceptual nature.

Indeed, the notions of a positive and of a negative factor are vague. In section 2.2.2 we have defined a positive factor as any event set defined by a predicate that meets requirements (a) to (d), while a negative factor is simply the complementary set of a positive factor. However, positive and negative factors do not constitute two distinct categories. Certain positive factors coincide with the complementary set of other positive factors. Hence, in certain cases there are multiple – positive and negative – factors available to typecast events. We shall not set up criteria that would bias such typecasts towards one of the two categories of factors. Pragmatic and epistemic criteria should decide in each single case whether an event is preferably taken to instantiate a positive or a negative factor. While these constraints may epistemically affect the application of SC in a number of cases, the conceptual accuracy of SC is not affected thereby.

3.8 *Abstraction from Relational Constraints*

Before we apply the thus far developed regularity theory to complex causal structures and address remaining problems, we have to pave the way for a simplification of our upcoming considerations. This chapter has shown that even the most modest minimal theories, i.e. minimal theories describing the simplest possible causal structures, are highly intricate first-order expressions. Ordinary causal structures, however, are of high complexity, involving a host of known and even more unknown factors. A first-order representation of such structures is ruled out by mere notational constraints. Due to their lack of transparency, full-blown first-order expressions of complex causal structures would not be serviceable to the goals pursued in the present study.

The first-order complexity of minimal theories is essentially due to the relational constraints causal structures impose on their factors and on the events that instantiate them. Causally related factors are instantiated by different events which, in turn, occur within the same spatiotemporal frame. Furthermore, instances of complex causes coincide. Expressing these constraints by means of the first-order formalism requires an extended technical apparatus. Yet, the decisive theoretical

⁶⁷ Dowe (2001), pp. 224-225, speaks of *epistemic blur* in this context.

work as regards the analysis of causal relevance proposed by a regularity account is done by the regularities subsisting among the factors involved in a corresponding causal structure. The mentioned relational constraints merely assure that a suitable subset of all the regularities holding between respective factors is chosen as a starting point for causal analyses. The relational constraints delineate the causally interesting regularities from causally meaningless regularities as “Whenever there is a table, there is a table”, “For every first human step on the lunar surface there is a first human non-stop balloon flight around the world”, or “Whenever there is a soccer game, there is a sport event”.

These relational constraints not only generate an exponential complexity increase of first-order minimal theories, they moreover are the reason for our resorting to predicate logic in the first place. If it were not for those constraints, a regularity account of causation could easily settle with ordinary propositional logic – minimal theories then being nothing but a material equivalence relationship among a minimally necessary disjunction of minimally sufficient conditions on the one hand and a purported effect factor on the other. Indeed, this propositional dependency relationship is exactly what is suggested by our abbreviated notation.

In order to simplify our upcoming discussions, we shall therefore completely settle with this abbreviated notation in the following. If the relational constraints causal structures impose are implicitly assumed to be satisfied by the factors contained in minimal theories and their instances, minimal theories expressed in our abbreviated notation can essentially be seen as propositional biconditionals. By explicitly formalizing these constraints in first-order logic as exemplified in this chapter, minimal theories expressed in our abbreviated notation should always be transferrable back into the first-order formalism. We thus take the following conditions to be implicitly satisfied from now on:

1. The factors in the antecedent of a minimal theory are instantiated by events that differ from the instances of the factor in the consequent of that theory.
2. The instances of the factors in the antecedent of a minimal theory and the instances of the consequent of that theory occur within the same spatiotemporal frame (such that it is clear what it means for events of respective factors to occur within the same spatiotemporal frame).
3. The factors in a minimally sufficient condition are coincidentally instantiated.

These three conditions taken jointly can be seen as the criterion that singles out the subset of regularities among factors that are of interest to causal analyses.

4. THE CHAIN-PROBLEM

4.1 Introduction

Up to recent years causal chains were given surprisingly little attention in studies on causation. Philosophical analyses of the causal relation focussed their interest on direct causal dependencies among causes and effects. Apparently, for a long time it was generally assumed that complex causal structures could be straightforwardly accounted for once a successful analysis of atomic causal dependencies would be available, i.e. of dependencies among single factors.¹ This confidence, however, has meanwhile turned out to have been premature for many prominent theoretical accounts of causation.

Lewis (1979), for instance, has pointed out that a counterfactual analysis, which in his view adequately accounts for atomic cause-effect relationships, is not capable of distinguishing between the first three graphs of figure 4.1 – provided that so-called *back-tracking* counterfactuals are taken to be well-behaved counterfactuals. For (a) and (b) the exact same counterfactual dependencies hold: $A \square \rightarrow B$, $A \square \rightarrow C$, $B \square \rightarrow A$, $B \square \rightarrow C$, $C \square \rightarrow A$, $C \square \rightarrow B$.² That is, these

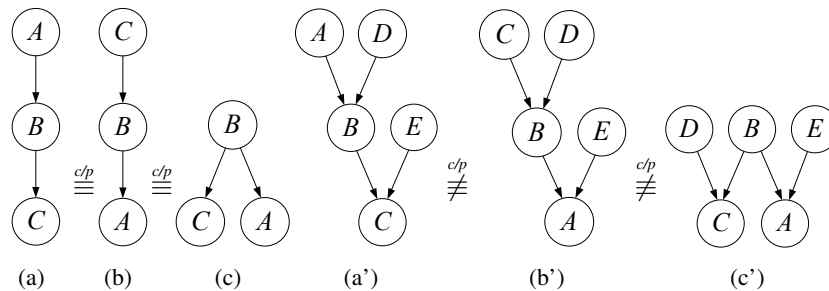


Fig. 4.1: (a), (b), and (c) represent three counterfactually and probabilistically equivalent causal structures – the respective equivalence being symbolized by “ \equiv_{cp} ”. Proper extensions of the three graphs yield the graphs (a’), (b’), and (c’), which no longer are counterfactually or probabilistically equivalent.

¹ One author who explicitly insisted on analyzing complex causal structures by resolving them into ‘atomic’ causal dependencies among single factors was e.g. Patrick Suppes in Suppes (1970). David Lewis embarked on a similar analytical strategy. He defined a causal relationship for exactly two factors by abstaining from complex structures, cf. Lewis (1973).

² As regards the notion of counterfactual dependence see Lewis (1973), pp. 560-561. In order to resolve such equivalencies Lewis rejects the adequacy of back-tracking interpretations of counterfactuals in Lewis (1979). He claims that counterfactuals have to be evaluated with respect to possible

structures are *counterfactually equivalent*. Hausman (1998) proposed to solve this problem by suitably extending the factor frame of these graphs.³ As soon as further alternative causes of the effects in (a), (b), and (c) are taken into consideration – as in (a'), (b'), and (c') – the epiphenomenal graph (c') can be distinguished by counterfactual means from the other two causal chains, which latter also become mutually discernable upon appropriate factor frame extensions. (b'), for instance, determines that $C \vee D \sqsupset B$. The same holds for graph (a'), but not for (c'). In contrast, structure (c') rules that $B \vee E \sqsupset A$, which is the case for (b') as well, but not for (a').

No later than Spirtes, Glymour, and Scheines (2000 (1993)) it has also become widely known in the field of probabilistic causation that building up complex structures from atomic cause-effect relations by means of probabilistic diagnostic algorithms is far more difficult than the pioneers of this account – e.g. Suppes (1970) – might have expected. Analyzing causal structures by means of probabilistic dependencies or independencies does not allow for a differentiation between the graphs (a), (b), or (c) in figure 4.1 either.⁴ For all of these graphs the same set of probabilistic (in-)dependencies holds: The conditional probabilities among the three factors yield dependencies except for A and C being independent given B in all three graphs, i.e. $P(A|BC) = P(A|B)$. Hence, the causal structures depicted in (a), (b), and (c) are *probabilistically equivalent*.⁵ That means, in light of only three analyzed factors, there is no way to distinguish between chains and epiphenomena by means of purely probabilistic techniques, i.e. without recourse to non-probabilistic relations as the direction of time – with the aid of which e.g. Suppes (1970) identifies chains and epiphenomena. Moreover, this ambiguity is independent of the quality of the available data; it persists even against the background of optimally significant data. As in the case of a counterfactual analysis, a purely probabilistic account of causation can only distinguish between chains and epiphenomena upon a suitable extension of the examined factor frame as conducted in the graphs (a'), (b'), and (c'). In graphs (a') and (c'), for instance, factors A and D are probabilistically independent, i.e. $P(AD) = P(A)P(D)$, while in (b') that independence does not hold. In contrast, the structures depicted in (b') and (c') induce a dependence of A and E , i.e. $P(A|E) > P(A)$, which is not the case for the structure represented by (a').

Counterfactual and probabilistic accounts of causation derive causal structures from sets of counterfactual and probabilistic (in-)dependencies, respectively. Yet,

worlds that are maximally similar to the actual world and that involve small *divergence miracles* just before the occurrence of the antecedent of a counterfactual. I join Hausman (1998), p. 114, in considering this resolution of counterfactual equivalencies as ad hoc and, thus, do not further discuss it here.

³ Cf. Hausman (1998), ch. 6.

⁴ Cf. e.g. Spirtes, Glymour, and Scheines (2000 (1993)), pp. 110, 169. In fact, Herbert Simon has already pointed out the problems a probabilistic account faces when it comes to a correct analysis of causal chains (cf. Simon (1954), pp. 468-469).

⁵ As regards the notion of probabilistic equivalence see Pearl (2000), pp. 19, 145, Verma and Pearl (1991), or Baumgartner and Graßhoff (2004), pp. 137-152.

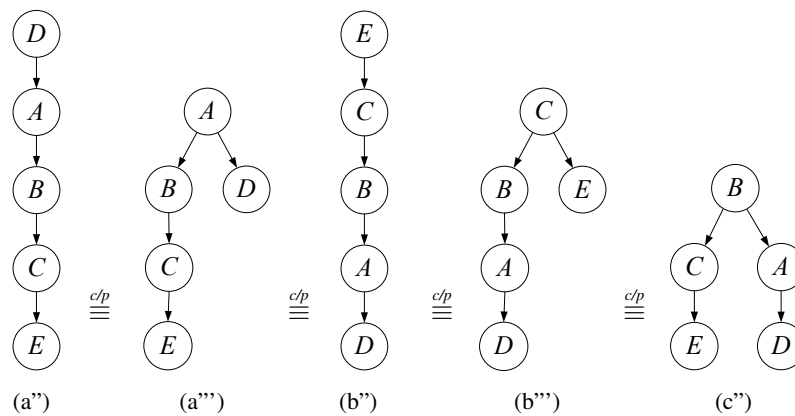


Fig. 4.2: If an extension of the factor frames of (a), (b), and (c) from fig. 4.1 yields either one of the five graphs depicted here, the counterfactual and probabilistic equivalencies among (a), (b), and (c) are not resolved. For (a), (b), (c) and the five structures in this figure the same counterfactual and probabilistic (in-)dependencies hold: All five factors are mutually counterfactually dependent and related in terms of the following probabilistic independencies: $P(A|BC) = P(A|B)$, $P(D|AB) = P(D|A)$ and $P(E|BC) = P(E|C)$.

the mapping of causal structures to such dependency sets is not generally unambiguous in case of complex structures, i.e. structures featuring more than two factors. Counterfactual and probabilistic accounts are recurrently forced to assign one and the same dependency set to more than one causal structure. A successful identification of causal structures thus often crucially depends on an adequate extension of the explored factor frame. The extensions of (a), (b), and (c) contained in figure 4.2 do not resolve the equivalencies among (a), (b), and (c). (a''), (a'''), (b''), (b'''), and (c'') are not to be kept apart within the frameworks of purely counterfactual or probabilistic accounts. Hence, it may – even relative to extended factor frames – happen that chains are not orientable or distinguishable from epiphenomena by counterfactual or probabilistic means.

The debate over complex causal structures, that has been conducted in the field of counterfactual and probabilistic accounts over the past years, has shown that an analysis of the causal relation – irrespective of a counterfactual or probabilistic background –, which starts with simplified processes among single factors and builds up complex structures modularly from this atomic inventory, is, contrary to primary intentions, forced to suitably extend factor frames when it comes to identifying causal chains. This, of course, raises the additional problem of specifying what ‘suitable extensions’ are as opposed ‘unsuitable’ ones.

Complexes of causal dependencies, however, are not a special but the ordinary form of causal structuring. The fact that accounts which analyze causal relevance in a modular way are forced to considerably adjust their original analytical strategy when confronted with complex structures indicates that the simple cause-effect relation among two factors might not be basal after all. Actually, it might be the other way around: Causal dependencies embedded in complex structures are the primary

analysans and dependencies between two single factors hinge on the relation of the factors in the pair to their causal neighbors. One central goal of this chapter will be to determine whether that hypothesis can be vindicated relative to a regularity theoretic background or not.

As against counterfactual and probabilistic accounts, a regularity framework operating with minimal theories focuses on more than two causally connected factors to begin with. As shown in the previous chapter, according to the here developed regularity account, the minimal complexity of causal structures involves at least two alternative causes for each effect. This focus not only allows for an unproblematic analysis of the direction of causation (cf. section 3.6.2), moreover, it is going to be of crucial importance to an identification of causal chains. MT thus has the simplest causal chain to be of the form of graphs (a') and (b'), not of (a) or (b).

Furthermore, as we shall see in the next chapter, the methods of causal reasoning induced by MT significantly diverge from comparable methods developed within the other theoretical frameworks. Whereas probabilistic diagnostic algorithms, for instance, evaluate probabilistic (in-)dependencies among pairs of factors and gradually build up complex structures from pairwise causal connections, regularity methodologies analyze causal dependencies among a principally open number of factors. The latter identify the place and function of a tested factor within an existing structure of arbitrary complexity by systematically varying the presence and absence of all known factors. This kind of coincidence analysis in combination with the definition of causal relevance for structures involving a smallest number of two alternative causes *prima facie* seems to provide the means to avoid the problems as regards an analysis of causal chains, which, in case of counterfactual and probabilistic accounts, are due to overly small factor frames. In the course of this chapter we shall, however, see that, contrary to first appearances, causal chains are not analyzed in such a forthright way by MT.

4.2 *Mapping Minimal Theories onto Complex Causal Structures*

The central question that needs to be answered when it comes to a representation of complex causal structures by minimal theories is this: Is every minimal theory assignable to exactly one causal structure or does it happen that the mapping of minimal theories onto complex structures is underdetermined as in the case of assigning sets of counterfactual and probabilistic (in-)dependencies to causal structures? Let us begin to consider this question – as it is usually done in the context of counterfactual and probabilistic accounts (cf. fig. 4.1) – by first looking at the simplest complex structures – structures involving three factors only. Depending on the grouping of the factors, the number of edges, and their orientation, three factors can be causally structured in several chain-shaped and epiphenomenal ways. Given two edges, for instance, 15 graphs can be generated: Three factors are groupable in threefold ways ($A - B - C$, $B - A - C$, and $A - C - B$) and for each of

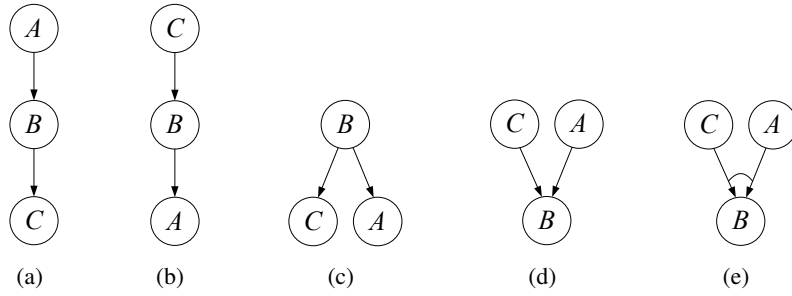


Fig. 4.3: The five graphs constructible based upon the skeleton $A - B - C$.

these so-called *graph-skeletons*⁶ four possible orientations of the edges exist, two chains, one epiphenomenon, and one structure of alternative causal relevance. Additionally, a fifth structure is constructible by means of an arch: a complex cause.⁷ Graphs consisting of three factors and two edges shall be labelled *3-2-graphs*. All of these 15 3-2-graphs represent a different causal structure. An analysis of causation must be able to mirror these differences. As the introductory remarks to this chapter have shown, without further extending factor frames this is not possible by means of counterfactual and probabilistic (in-)dependencies alone. We now have to evaluate how a regularity account handles this problem. Each of the different 3-2-graphs needs to be assigned to a different minimal theory.

For reasons of brevity, we will not assign a minimal theory to each of the 15 3-2-graphs. We shall focus on one skeleton only: $A - B - C$. Minimal theories are analogously assignable to the other 10 3-2-graphs. Figure 4.3 exhibits the five causal graphs based on the skeleton $A - B - C$. What minimal theories represent the causal structures depicted by graphs (a), (b), (c), (d), and (e)?

These graphs, as most causal graphs, are incomplete. The antecedents of minimal theories, however, feature necessary conditions for their consequents. As graphs (a), (b), (c), (d), and (e) neither specify any factors that complete minimally sufficient conditions nor (further) alternative causes of their respective effects, we resort to variables X_1 and X_2 on the one hand and to Y_A , Y_B , and Y_C on the other.⁸ In this manner, the implicit incompleteness of (a), (b), (c), (d), and (e) can be rendered explicit. The following complex minimal theories represent the causal structures depicted by the graphs of figure 4.3:

$$\begin{aligned}
 (AX_1 \vee Y_B \Rightarrow B) \wedge (BX_2 \vee Y_C \Rightarrow C) & \quad (MT_a) \\
 (CX_1 \vee Y_B \Rightarrow B) \wedge (BX_2 \vee Y_A \Rightarrow A) & \quad (MT_b) \\
 (BX_1 \vee Y_A \Rightarrow A) \wedge (BX_2 \vee Y_C \Rightarrow C) & \quad (MT_c) \\
 AX_1 \vee CX_2 \vee Y_B \Rightarrow B & \quad (MT_d) \\
 ACX_1 \vee Y_B \Rightarrow B & \quad (MT_e)
 \end{aligned}$$

⁶ As regards the notion of a *graph-skeleton* see Pearl (2000), p. 12.

⁷ For details on the graphical notation or complex causes see section 2.3.4.

⁸ Cf. section 3.5.

There are no logical equivalencies among these minimal theories. Therefore, each of them is straightforwardly assignable to one of the graphs in figure 4.3. Assuming that the factors contained in MT_a to MT_e remain part of a respective minimal theory across all factor frame extensions, each of these theories explicates one of the causal statements expressed by a corresponding causal graph. This contention will need to be substantiated in the following. Thus, we now set out to show that, contrary to the representation of 3-2-graphs by sets of counterfactual or probabilistic dependencies, the assignment of causally interpretable minimal theories to 3-2-graphs is not underdetermined.

To this end we first have to determine the conditions under which two double-conditional expressions⁹ state the same causal structure, i.e. correspond to the same causally interpretable minimal theory. Thus, a notion of equivalence for causally interpretable minimal theories needs to be developed. Since minimal theories are only causally interpretable provided that their constituent factors remain part of them across all factor frame extensions, causally interpretable minimal theories correspond to semi-formal expressions – first-order expressions supplemented by the non-formalizable condition assuring the extendability of factor frames which does justice to (PPR).¹⁰ Due to this semi-formality, common notions of logical equivalence cannot be resorted to when it comes to determining the criteria under which minimal theories stipulate identical causal structures.

Minimal theories assess which coincidences of the factors they are constituted of occur and which ones do not. The factors contained in a minimal theory Φ correspond to the factor frame \mathcal{F} of Φ .¹¹ Provided that no extension of \mathcal{F} renders a factor $Z_1 \in \mathcal{F}$ redundant, Φ is causally interpretable. If that is the case, the behavior of the factors in \mathcal{F} is regulated by a causal structure which is represented by Φ . A causal structure can be said to regulate the behavior of the factors in \mathcal{F} if not all of the 2^n logically possible coincidences among the n factors in \mathcal{F} are empirically possible. A coincidence X shall be labelled *empirically possible* iff there exists at least one instance of X in the past, present, or future. The causal structure depicted by graph (a), for instance, determines that whenever the complex cause of B is instantiated, which A is part of, there is an instance of B . Hence, a coincident instantiation of AX_1 and \bar{B} is empirically impossible according to graph (a). This regulating of the behavior of the factors in a given frame is the characteristic feature of causal structures that is mirrored by minimal theories. As causal structures, minimal theories are not compatible with all 2^n logically possible coincidences among the n factors in their frame. They are not tautologous expressions. In accordance with graph (a), MT_a does not allow for a coincident instantiation of the factors AX_1 and \bar{B} . Whenever AX_1 is instantiated, MT_a assesses that B is instantiated as well.¹² In this sense, a minimal theory Φ selects a proper subset

⁹ Cf. section 3.5 above, p. 103.

¹⁰ Cf. section 3.6.1.

¹¹ Cf. section 2.2.2, p. 64.

¹² Bear in mind that \bar{B} in our abbreviated notation is to be read as a negative existential statement claiming that no event of type B is instantiated (cf. section 3.6.4).

of the 2^n logically possible coincidences among the factors in its frame, such that the coincidences in that subset are maintained to be empirically possible by Φ , in which case the coincidences in that subset are said to be *compatible* with Φ .

Against this background, two minimal theories Φ and Ψ shall be seen as equivalent in causal respects if they have a common factor frame \mathcal{F} , such that no factor $Z_1 \in \mathcal{F}$ is rendered redundant by extensions of \mathcal{F} , and Φ is compatible with a coincidence X of the factors in \mathcal{F} iff Ψ is thus compatible. For convenience, we shall refer to a factor frame \mathcal{F} whose factors are not rendered redundant by arbitrary extensions of \mathcal{F} as a *persistent factor frame*. Furthermore, the set of coincidences compatible with a given minimal theory Φ shall be labelled the *coincidence frame* of Φ within the factor frame \mathcal{F} of Φ .¹³ Hence, two minimal theories Φ and Ψ with persistent factor frames are causally equivalent iff Φ and Ψ share a common factor frame and therein a common coincidence frame. That means that causally equivalent minimal theories regulate the behavior of the factors contained in their common frame in exactly the same way. In order to terminologically mark this notion of equivalence off against other notions of equivalence, I shall in this context speak of *mt-equivalence*.

Persistent factor frame: The factor frame \mathcal{F} of a minimal theory Φ (of a causal structure or causal graph G , respectively) is called *persistent* iff no factor $Z_1 \in \mathcal{F}$ is rendered redundant by arbitrary extensions of \mathcal{F} .

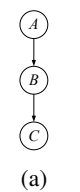
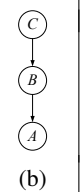
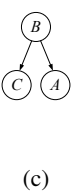

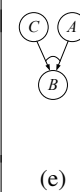
Coincidence frame: The coincidence frame of a minimal theory Φ (of a causal structure or causal graph G , respectively) within a given factor frame \mathcal{F} is the set of coincidences in \mathcal{F} compatible with Φ (with G).

MT-equivalence: Two minimal theories Φ and Ψ with persistent factor frames are *mt-equivalent* iff Φ and Ψ share both a common factor frame and a common coincidence frame.

The notion of *mt-equivalence* shall now be applied to the theories MT_a to MT_e . The factor frame of the graphs in figure 4.3 consists of the three factors A , B , and C , which can be instantiated in 8 logically possible combinations. The incompleteness of the graphs in 4.3 additionally calls for introducing a number of unknown factors, which, as mentioned above, constitute the domain of the variables X_1 and X_2 on the one hand and Y_A, Y_B, Y_C on the other. Combinatorially assembling coincidences from the inventory of this factor frame yields $2^8 = 256$ logically possible coincidences, against the background of which MT_a to MT_e would have to be compared with respect to compatibilities and thus *mt-equivalencies*. In order to simplify matters, table 4.1 picks out those logically possible coincidences that are particularly interesting when it comes to checking for potential *mt-equivalencies*. Table 4.1 selects coincidences such that unknown factors X_1 and X_2 in minimally sufficient conditions containing A , B , or C are instantiated while

¹³ The notions of a *coincidence* and a *factor frame* will be applied analogously to causal structures and graphs below.

	MT_a	MT_b	MT_c	MT_d	MT_e
$ABC X$	1	1	1	1	1
$AB\bar{C} X$	0	1	0	1	1
$A\bar{B}C X$	0	0	1	0	0
$\bar{A}BC X$	1	0	0	1	1
$\bar{A}B\bar{C} X$	0	1	1	0	1
$\bar{A}\bar{B}C X$	1	0	1	0	1
$\bar{A}\bar{B}\bar{C} X$	0	0	0	1	1
$\overline{ABC} X$	1	1	1	1	1

				
(a)	(b)	(c)	(d)	(e)

Tab. 4.1: This table assigns the complex minimal theories indicated in the top line to the causal structures listed in the bottom line. The list of compatibilities with logically possible coincidences within the frame consisting of A , B , and C serves as the criterion according to which minimal theories are mapped onto causal graphs. That means, minimal theories are mapped onto causal structures depending on their coincidence frames. X represents X_1X_2 .

A , B , and C systematically vary as to their presence and absence. The conjunction of X_1 and X_2 is simply represented by X in table 4.1, i.e. $X = X_1X_2$. In contrast, the presence and absence of the disjuncts in the domain of Y_A , Y_B , and Y_C is left unspecified. On every line of table 4.1 Y_A , Y_B , and Y_C can be instantiated or not.

Each of the selected coincidences is marked to be compatible or incompatible with a corresponding minimal theory by table 4.1. If the former is the case, the pertaining field contains a “1”, if a coincidence is incompatible with a given minimal theory, the pertaining fields are marked by “0”. This tabular compilation shows that none of the theories MT_a to MT_e have a common coincidence frame. None of these minimal theories are thus *mt*-equivalent.

By means of a respective coincidence frame table 4.1 assigns MT_a , MT_b , MT_c , MT_d , and MT_e each to exactly one causal graph of figure 4.3. Hence, the causal structures depicted in figure 4.3 are represented by minimal theories without further adjusting or amending the primary analysans of our account of causation. The reason for this – in comparison to counterfactual or probabilistic accounts – plain handling of 3-2-graphs lies in the fact that the regularity account developed here focuses on complex structures to begin with. MT does not analyze causal relevance on the basis of artificially simplified structures, such as pairs of causally connected factors, but meets the complexity of causal structures from the outset.

4.3 Chains vs. Entangled Ephiphenomena

4.3.1 Inference to Complex Causal Structures – The Basic Idea of an Inference Procedure

Provided that the sentence

(P) Different causal structures generate different coincidence frames.

holds generally and not only in the case of the examples in figure 4.3, the previous section has laid down the foundation of a procedure that derives causal structures from sets of coincidences. For if (P) holds generally, coincidence frames can be resorted to as a criterion that individuates complex causal structures. The notion of *mt*-equivalence makes sure that causal structures that generate different coincidence frames do not correspond to *mt*-equivalent minimal theories. Hence, if (P) is valid, an unambiguous mapping of minimal theories onto causal structures is possible. Against this background, complex causal structures could unequivocally be inferred from coincidence frames. Or put differently: Coincidence frames would – given the validity of (P) – serve as empirical data based on which minimal theories and thereby, in a second step, causal structures could be derived. Of course, all details of such an inference procedure are yet to be developed.

4.3.2 Different Causal Structures, one Coincidence Frame

If (P) is valid, every causal structure – irrespective of its complexity – regulates the behavior of its factors in a characteristic and unique way, by means of which it could be identified. It all hinges on the validity of (P). How this validity is to be evaluated is clear: We have to search for two or more *different* causal structures that do *not differ* with respect to their coincidence frames. If such structures exist, (P) is false, if not, (P) is true. The result of this search can be anticipated at this point: There in fact exist different causal structures that generate identical coincidence frames. Thus, (P) is false. Figure 4.4 exhibits two structures differing in causal respects – graph (h) represents a chain, graph (i) an ephiphenomenon –, yet sharing

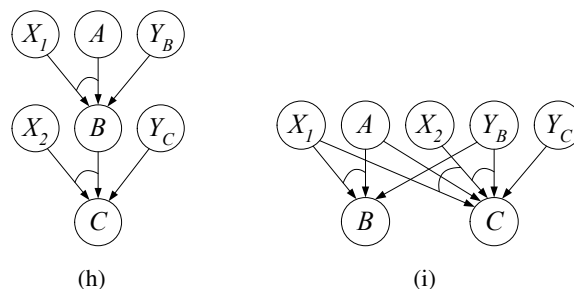


Fig. 4.4: Two different causal structures that share a common coincidence frame. The structures have been completed by use of variables.

a common coincidence frame. Both structures regulate the behavior of their factors in an identical manner.

Such as to prove that the coincidence frames of (h) and (i) in fact coincide, we could go through all of the logically possible combinations of the 7 factors and variables that are involved in these graphs and thereby show that any of these $2^7 = 128$ coincidences is compatible with one of the two graphs iff it is also compatible with the other. As in case of the graphs (a) to (e), however, listing 128 coincidences is neither transparent nor possible here. Hence, the identity of the coincidence frames of (h) and (i) will be demonstrated indirectly in the following. To this end we shall first assign minimal theories to the two graphs and then show that these minimal theories are *mt*-equivalent.

The minimal theories that correspond to the graphs (h) and (i) are:¹⁴

$$\begin{aligned} (AX_1 \vee Y_B \Rightarrow B) \wedge (BX_2 \vee Y_C \Rightarrow C) & \quad (MT_h) \\ (AX_1 \vee Y_B \Rightarrow B) \wedge (AX_1X_2 \vee Y_BX_2 \vee Y_C \Rightarrow C) & \quad (MT_i) \end{aligned}$$

The following consideration shows that these complex minimal theories are *mt*-equivalent. They consist of two conjunctively connected simple theories, the first of which appears both in MT_h and MT_i . This first conjunct of MT_h and MT_i stipulates:

(Q) Whenever $AX_1 \vee Y_B$ is instantiated, there is an instance of B , and whenever there is an event of type B , $AX_1 \vee Y_B$ is instantiated.¹⁵

Clearly, every logically possible coincidence within the factor frame under consideration, which is incompatible with (Q), is incompatible both with MT_h and MT_i , for they both assert (Q). Therefore, a coincidence K that could establish a difference between the coincidence frames of MT_h and MT_i must be compatible with the second conjunct of either MT_h or MT_i while it is incompatible with the second conjunct of the other theory.

Such a coincidence K would have to satisfy either of the following conditions:

- in K , $BX_2 \vee Y_C$ is not instantiated, while C and $AX_1X_2 \vee Y_BX_2 \vee Y_C$ both are,
- in K , $BX_2 \vee Y_C$ is instantiated, but there is neither an instance of C nor of $AX_1X_2 \vee Y_BX_2 \vee Y_C$,
- in K , there is no instance of $AX_1X_2 \vee Y_BX_2 \vee Y_C$, but one of C and $BX_2 \vee Y_C$,
- in K , there is an instance of $AX_1X_2 \vee Y_BX_2 \vee Y_C$, but neither one of C nor one of $BX_2 \vee Y_C$.

Yet, every coincidence that satisfies any of these conditions, violates the constraints imposed by (Q) and is thus incompatible with both MT_h and MT_i . For (Q) implies that for every instance of a disjunct of $AX_1X_2 \vee Y_BX_2 \vee Y_C$ there is an instance of

¹⁴ For details on the abbreviated notation – especially in case of Y_BX_2 in MT_i – implemented here cf. section 3.5, p. 102.

¹⁵ As mentioned in section 3.8, relational constraints are not explicitly appealed to any more in this chapter. They are implicitly assumed to be satisfied.

a disjunct of $BX_2 \vee Y_C$ and that for every instance of a disjunct of $BX_2 \vee Y_C$ there is an instance of a disjunct of $AX_1X_2 \vee Y_BX_2 \vee Y_C$. Therefore, there cannot be a coincidence which is compatible or incompatible with *one* of the minimal theories under discussion only. This proves that MT_h and MT_i are *mt*-equivalent. Any logically possible coincidence within the factor frame of MT_h and MT_i is either compatible with both theories or with none of them.

This argument can analogously be put forward with respect to the graphs (h) and (i). The second layer¹⁶ of (h) and the left subgraph of (i), i.e. the subgraph representing the causal structure behind the behavior of B , are identical. Hence, for a coincidence to constitute a difference between these two graphs as a whole it would have to be compatible with the first layer of (h) only or the right subgraph of (i) only. However, any coincidence satisfying this condition violates both the second layer of (h) and the left subgraph of (i) and is thus incompatible with both causal structures in figure 4.4. Therefore, these structures have the same coincidence frame.

Notwithstanding the fact that both causal structures regulate the behavior of the involved factors identically, they differ in causal respects. (h) and MT_h determine that A is *indirectly* causally relevant for C , while the same factor A is attributed *direct* causal relevance for C by (i) and MT_i . Furthermore, B is causally relevant for C in (h), yet not in (i). (h) represents a causal chain, (i) an epiphenomenon. However, all of these significant differences in causal structuring do not affect the behavior of the involved factors. Any coincidence is either compatible with both (h) and (i) or with neither of the two graphs.

(h) and (i) are by far not the only graphs describing different causal structures with identical coincidence frames. To every causal graph G_1 consisting of at least two layers there exists an *mt*-equivalent graph G_2 with one layer only, i.e. a graph such that there is no vertex being both tail and head in G_2 .¹⁷ Graphs of type G_1 have been labelled *chain-graphs* in section 3.6.6, graphs of type G_2 shall correspondingly now be referred to as *epi-graphs*. By means of the following procedure every chain-graph consisting of two layers is transferrable into an epi-graph with an identical coincidence frame.

- (1) Detach the first layer of a two-layer chain-graph G from its second layer.
- (2) Remove those vertices from the first layer of G that are tails of the second layer, including the paths originating from them.
- (3) Connect the heads H_2 of the second layer of G to the tails T_1 of the first layer of G such that trails¹⁸ originally including the tails of the second and the heads of the first layer (T_2/H_1), $H_2 \longrightarrow T_2/H_1 \longrightarrow T_1$, are replaced by corresponding paths $H_2 \longrightarrow T_1$.

¹⁶ Cf. section 3.6.6, p. 137.

¹⁷ For details on the graph theoretical terminology see section 2.3.2.

¹⁸ The notion of a *trail* has been introduced in section 3.6.6 above.

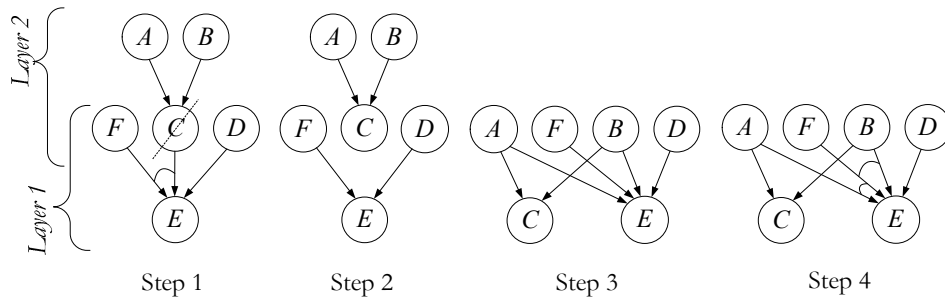


Fig. 4.5: By means of this procedure every chain-graph containing two layers is transferrable to an *mt*-equivalent epi-graph.

- (4) Join the edges of the first layer, that have originally been connected by an arch with an edge $T_2/H_1 \rightarrow T_1$, by means of an arch with every path newly generated in step (3).

Figure 4.5 lists the four steps of this transference procedure. It can be applied recursively to chain-graphs consisting of an arbitrary number of layers. In such a manner, any many-layer chain-graph can be transferred to a one-layer epi-graph with an identical coincidence frame. Figure 4.6 illustrates the reduction of a three-layer to a one-layer graph. As graphs represent causal structures, all of this shows that to every causal chain there exists an epiphenomenon with an identical coincidence frame. Every causal chain is reducible to an epiphenomenal structure.

Epiphenomena with a coincidence frame matching the coincidence frame of a causal chain have a very specific structure: *All* factors that are part of a minimally sufficient condition of one effect – e.g. *B* in graph (i) – are also contained in the minimally sufficient conditions of the other effect – *C* in case of (i). Two factors satisfying this structural pattern such as *B* and *C* shall be referred to as *entangled* factors. Entangled factors are not only contained in epiphenomena as (i), but

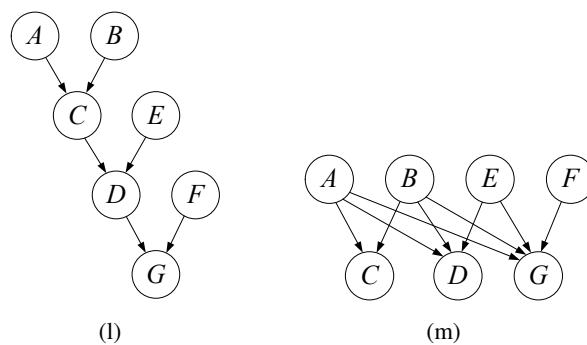


Fig. 4.6: In this example a chain-graph consisting of three layers is reduced to an *mt*-equivalent epi-graph. For reasons of simplicity, (i) and (m) are assumed to be complete with respect to causally relevant factors for *C*, *D*, and *G*.

notably in chains. B and C are entangled in graph (h) as well. Two factors are entangled iff every conjunct of a minimally sufficient condition of one of the two factors is also a conjunct of at least one minimally sufficient condition of the other factor.

Every epiphenomenon that shares its coincidence frame with a causal chain features at least two entangled factors. That means, at most one effect of such an epiphenomenon has causally relevant factors that are not part of a minimally sufficient condition of the other effect. For the purpose of an easy reference to such epiphenomenal structures, I shall in this context speak of *entangled epiphenomena*.

Entangled factors: Two factors A and B are entangled iff all factors contained in minimally sufficient conditions of A are part of a minimally sufficient condition of B or all factors contained in minimally sufficient conditions of B are part of a minimally sufficient condition of A .

Entangled epiphenomenon: An epiphenomenon is called *entangled* iff its effects are entangled factors.

There thus exists an entangled epiphenomenon to every causal chain such that the epiphenomenon and the chain have a common coincidence frame. Correspondingly, to every causally interpretable minimal theory representing a chain there exists an *mt*-equivalent minimal theory representing an entangled epiphenomenon. The converse, however, does not hold. Table 4.1 has shown that not to every epiphenomenon there exists a chain with a common coincidence frame. The epiphenomenon (c), for instance, which is not entangled, is easily distinguishable from the chains (a) and (b). The causal structure behind the behavior of two factors A and B cannot be a chain if the following conditions hold: Among the minimally sufficient conditions of A there is a condition containing at least one factor Z_1 which is not part of any minimally sufficient condition of B ; and among the minimally sufficient conditions of B there is a condition containing at least one factor Z_2 which is not part of any minimally sufficient condition of A . If A and B , furthermore, have at least one minimally sufficient condition in common, their behavior must be regulated by an epiphenomenal structure. All of these ‘ordinary’ epiphenomena are identifiable via their coincidence frames.

As long as causal inference methodologies exclusively analyze coincidence frames – or sets of coincidences –, an unambiguous inference to causal chains is excluded in principle. Based on such pure coincidence analyses every causal process which is commonly assumed to be structured in terms of a chain could just as well be modeled as an entangled epiphenomenon. This, of course, is a finding that heavily conflicts with common intuitions. The fact that causes and effects are ordinarily concatenated in chains is one of the core features of our understanding of the causal relation. The inability of pure coincidence analyses to distinguish between chains and epiphenomena thus is a serious problem. I call it the *chain-problem*.

Chain-problem: Whenever, within a factor frame \mathcal{F} , there is a set R of occurring coincidences such that R is compatible with a causal chain, R is compatible with an entangled epiphenomenon as well. By means of pure coincidence analyses it is thus impossible to determine whether R stems from a causal chain or an epiphenomenon.

4.4 Causal Intuition and the Distinction Between Chains and Entangled Epiphenomena

Let us look at a concrete example that illustrates the impossibility to distinguish chains and entangled epiphenomena simply based on sets of coincidences occurring within a given factor frame. Suppose, within a factor frame \mathcal{F}_1 consisting of the factors A , B , C , D , and E , exactly the following 8 coincidences occur:

$$\begin{array}{l} ABCDE \\ ABC\bar{D}\bar{E} \\ A\bar{B}CDE \\ A\bar{B}C\bar{D}\bar{E} \\ \bar{A}BCDE \\ \bar{A}BC\bar{D}\bar{E} \\ \bar{A}\bar{B}CDE \\ \bar{A}\bar{B}C\bar{D}\bar{E} \end{array} \quad (R)$$

For reasons of transparency, this example is thus designed that the behavior of the factors in R is regulated by simple dependencies, that overall satisfy the conditions as to the minimal complexity of causal structures,¹⁹ such that no conjectures with respect to the existence of unknown causal factors are called for. That is to say, in R the following dependencies hold: A and B are each minimally sufficient for C ; A , B , C , and D are each minimally sufficient for E . This follows from the fact that there are no coincidences including $A\bar{C}$, $B\bar{C}$, $A\bar{E}$, $B\bar{E}$, $C\bar{E}$, and $D\bar{E}$ in R , i.e. A is never instantiated without an instance of C or E . The same holds for B . Likewise, C and D are never instantiated without an instance of E . Furthermore, the disjunction $A \vee B$ is minimally necessary for C , for whenever C is instantiated, there is an instance of A or of B . The necessary condition of E , however, cannot be unambiguously minimalized, because three of the minimally sufficient conditions of E , namely A , B , and C , never demonstrate their potential causal relevance for E independently of each other. Whenever A or B are instantiated, there is an instance of C as well.

The coincidences in R are therefore compatible with two *mt*-equivalent complex minimal theories:

$$\begin{array}{ll} (A \vee B \Rightarrow C) \wedge (C \vee D \Rightarrow E) & (MT_n) \\ (A \vee B \Rightarrow C) \wedge (A \vee B \vee D \Rightarrow E) & (MT_o) \end{array}$$

¹⁹ Cf. section 3.6.2, p. 113.

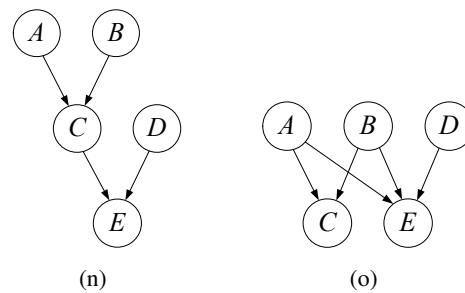


Fig. 4.7: A causal chain and an epiphenomenon that both could underly the behavior of the factors in \mathcal{F}_1 .

These minimal theories – in sequential order – represent the graphs in figure 4.7. Even though MT_n and MT_o are *mt*-equivalent they are opposed as regards the causal structures they exhibit. Causal relevance of C for E ensues from MT_n , yet not from MT_o . A regularity account of causation as developed thus far, hence, is drawn towards both attributing and not attributing causal relevance for E to C . As long as the causal relevance of factors is analyzed via their membership in the antecedents of minimal theories, it is not to be avoided that either MT_n or MT_o is denied the status of a minimal theory. Inextricably linked with this problem is the question as to based on what rules a coincidence frame as R should be taken to warrant an inference to MT_n or MT_o and graphs (n) or (o). The fact that this question can be raised may seem surprising, for in view of concrete interpretations of the factors in \mathcal{F}_1 , causal intuitions are as determined as can be when it comes to opting for the chain or the epiphenomenon given a coincidence frame as R .

Interpretation (I): Assume a car engine can be started in two ways only: either by turning the key in the starter lock or by short-circuiting the ignition cable. Whenever the engine is running, the corresponding car begins to move. The car can be set in motion by alternative factors also, such as towing or pushing, i.e. by external impulses:

- A = Turning the key in the starter lock
- B = Short-circuiting the ignition cable
- C = Running engine
- D = External impulse
- E = Motion of the car.

Interpretation (I) clearly suggests the underlying causal structure to be (n). Hence, provided that the factors in R are interpreted according to (I), we tend to model the underlying process in terms of a chain.

Interpretation (II): Suppose in a particular city there are exactly two power stations. The power supply of a specific house, say house r , in that city entirely depends on the power production in at least one of the two stations. Another

house, call it s , is equipped with a generator for cases of citywide power failures. Whenever one of the two power stations produces electricity, both r and s are power supplied:

- A = Power production by station 1
- B = Power production by station 2
- C = Power supply of house r
- D = Power production by the generator in s
- E = Power supply of house s .

If the coincidences in R are interpreted in the vein of (II), the behavior of the factors in \mathcal{F}_1 is intuitively seen to be regulated by an epiphenomenal structure of type (o).

Intuitively there is no doubt that the causal process starting with turning the key in the starter lock and resulting in the motion of the car has the form of a chain and that there is an epiphenomenal structure behind the power supply of the houses r and s . Still, the coincidences in R alone neither warrant the first nor the second of these causal inferences. Therefore, our firm intuitions as regards causal modeling relative to a respective interpretation cannot exclusively be based on the coincidence frame R . How do we intuitively distinguish between chains and entangled epiphenomena when confronted with specific interpretations of the factors in a given frame? This is one of the prevalent questions raised by the chain-problem.

4.5 *Inference to Chains and the Conceptual Fundament of a Regularity Theory*

The chain-problem both affects causal reasoning and the conceptual fundament of the thus far developed regularity theory. Or put differently, there are two versions or aspects of the chain-problem: an epistemic and a conceptual one. With an epistemic emphasis, the chain-problem can be formulated thus: Whenever a set of coincidences is causally analyzed and it is found that this set is compatible with a chain, a causal inference to a chain is equally warranted as an inference to an entangled epiphenomenon. In contrast, conceptually emphasized the chain-problem can be put as follows: Complex minimal theories are not always unambiguously assignable to complex causal structures; to every minimal theory representing a chain there exists an *mt*-equivalent – yet causally contrary – minimal theory representing an entangled epiphenomenon.

These two aspects of the chain-problem have twofold consequences: First, the epistemic version shows that methodologies of causal inference cannot exclusively rely on coincidence frames. Second, assuming that causal relevance is to be analyzed – in accordance with MT – by means of the membership of factors in antecedents of minimal theories with persistent factor frames, not all extension resistant double-conditional expressions can be granted the status of causally interpretable minimal theories. One of e.g. MT_n and MT_o has to be denied that status.

It must be excluded that the conceptual fundament of our theory forces us to both attribute and not attribute causal relevance to a factor at the same time. We shall see that the central notion that needs to be adapted such as to resolve these ambiguities is the notion of a complex minimal theory.

Chains are one of the most fundamental causal structures. Neither a methodology of causal reasoning nor a theoretical account of causation can be considered valuable as long as they suffer from the chain-problem. This problem thus imperatively needs to be solved. The above considerations indicate that such a solution must be twofold. On the one hand, we need to determine what causal reasoning has to resort to besides coincidence frames, and, on the other hand, the notion of a causally interpretable minimal theory needs to be accommodated.

4.5.1 Causal Interpretation of Entangled Factors

Such as to determine what has to be added to mere coincidence data in order to unambiguously identify causal chains, it will be useful to consider how we intuitively chose the chainlike and the epiphenomenal structures in case of interpretations (I) and (II). In both exemplary interpretations an utterly simplistic causal structure is assumed to regulate the behavior of the factors in \mathcal{F}_1 . Thus, both examples are ‘unnaturally’ constructed for the sake of simplicity. Real cars can be started in manifold ways, each of which involve many more factors than merely the turning of a starter key or the short-circuiting of a cable. The same holds for power supplied houses. The mere operating of power stations is commonly neither sufficient nor necessary for the power supply in houses. It takes a sophisticated infrastructure to bring the electricity to the consumers and generators or batteries can supply any houses, not just s, with power.

Causal intuitions, however, do not evaluate the causal structure behind moving cars or power supplied houses based on simplistic backgrounds as given in (I) and (II). In a way thus, causal intuitions ignore or dismiss the construction of our examples. We intuitively analyze these examples against a much more complex background, within a much extended factor frame. Here, as will be shown promptly, lies the key to a solution of the chain-problem in its epistemic variety. We unhesitatingly prefer the causal chain in view of interpretation (I) and the epiphenomenon in light of interpretation (II), because we do not judge these cases relative to a restricted factor frame as \mathcal{F}_1 .

\mathcal{F}_1 is extendable in manifold ways when it comes to generating a more realistic scenario with respect to the car example. An electrical impulse is transmitted from the turning of the starter lock to the spark plug by means of adequate wiring. After a firing of the spark plug the engine starts such that its kinetic energy is transmitted to the wheels through the axis. Then the car begins to move. Hence, factors can be introduced into \mathcal{F}_1 that mediate the causal influence of the turning of the starter key to the movement of the car. Or there are many other ways to trigger the spark plug, for instance by use of jumper cables or of a mechanical starting device. If the factors in \mathcal{F}_1 are interpreted as in (I), every extension of the original (simplistic)

factor frame will retain the entanglement of C and E . Every newly introduced factor or conjunction of factors that is part of a minimally sufficient condition for the running of the car engine will also be part of a minimally sufficient condition for the car movement. This holds because C is itself part of a minimally sufficient condition of E . Assume this minimally sufficient condition to be CX_1 . Every substitution of C in CX_1 by a minimally sufficient condition of C results in a minimally sufficient condition of E , for CX_1 is minimally sufficient for E .

In an analogous way, \mathcal{F}_1 is extendable when interpreted along the lines of (II). The power produced at the two stations is transferred to the consumers by means of suitable supply lines, electrical power can be imported or generators or batteries can be hooked up to any houses, not just s . The latter is the main reason why causal intuitions opt for the epiphenomenon in light of interpretation (II). We know that generators or batteries can be installed in any house such that the corresponding house is power supplied *independently* of any other houses. When evaluating the causal structure behind the power station example we intuitively analyze a factor frame that not only includes the power production by a generator in house s , but also a factor such as “power production by a generator in r ”. We thus extend \mathcal{F}_1 by at least one alternative cause of C – one alternative cause that is *independent* of the minimally sufficient conditions of E . Relative to such an extension there will be cases in which house r is power supplied while house s is not, for instance, when neither of the two power stations operate, the generator in s is out of order, yet house r is supplied with electricity by its own generator. Accordingly, C will not be sufficient for E any longer. Hence, in case of interpretation (II) extending \mathcal{F}_1 supplements R by additional coincidences to the effect that the entanglement of C and E is suspended. This yields a coincidence frame that is unambiguously assignable to an (ordinary) epiphenomenon.

Even though both the car and the power station example are thus designed that they are assumed to be underlain by two different causal structures that generate identical coincidence frames, extensions of \mathcal{F}_1 lead to important differences depending on whether \mathcal{F}_1 is interpreted in terms of (I) or of (II). In order to clearly bring out those differences, let us concisely summarize the above considerations. Against the background of interpretation (I), \mathcal{F}_1 is extendable by, say, the factors:

- G = Starting the engine by use of jumper cables
- H = Starting the engine by use of a mechanical starting device
- I = Transmission of an electrical impulse to the spark plug.

The thus extended factor frame shall be labelled \mathcal{F}_2 . In contrast to interpretation (I), \mathcal{F}_1 , when interpreted according to (II), is e.g. extended by the following factor:

- L = Power production by the generator in r .

This expansion of \mathcal{F}_1 , interpreted in terms of (II), will be referred to as \mathcal{F}_3 .

The expansions of \mathcal{F}_1 to \mathcal{F}_2 and \mathcal{F}_3 differ in an important respect: The expansion of \mathcal{F}_1 to \mathcal{F}_2 is *structure-conserving*, whereas the same does not hold for the

expansion of \mathcal{F}_1 to \mathcal{F}_3 . The feature of the graphs (n) and (o), which is characteristic for the chain-problem, is the entanglement of C and E . This entanglement is conserved when \mathcal{F}_1 – interpreted along the lines of (I) – is extended to \mathcal{F}_2 . It is not only conserved upon extending \mathcal{F}_1 by G , H , and I . Every factor that is newly introduced into a chainlike structure as (n) and that is part of a minimally sufficient condition of C , will be part of a minimally sufficient condition of E as well. This follows from the mere structure of causal chains. Consider anew the minimal theory MT_n which represents graph (n) of figure 4.7. Suppose, the factor frame of MT_n is extended and – among other things – additional factors are introduced into the set of minimally sufficient conditions of C , e.g. as follows:

$$(AX_1 \vee BX_2 \vee X_3 \Rightarrow C) \wedge (CX_4 \vee DX_5 \Rightarrow E) \quad (MT'_n)$$

From the mere form of MT'_n in combination with the assumed transitivity of \mathbf{R} ²⁰ it follows that whatever factor is part of a minimally sufficient condition of C must be part of a minimally sufficient condition of E as well. Or formally:

$$\begin{aligned} (AX_1 \vee BX_2 \vee X_3 \Rightarrow C) \wedge (CX_4 \vee DX_5 \Rightarrow E) \wedge \forall x, y, z (\mathbf{R}xy \wedge \mathbf{R}yz \rightarrow \mathbf{R}xz) \\ \vdash AX_1 X_4 \mapsto E \\ \vdash BX_2 X_4 \mapsto E \\ \vdash X_3 X_4 \mapsto E. \end{aligned}$$

If we abstract from relational constraints and treat a minimal theory on par with a propositional biconditional, this is straightforwardly proven by means of the substitutability of equivalents (SUB&: $A \leftrightarrow B \vdash A \wedge C \leftrightarrow B \wedge C$) and distributive laws (DIS: $A \wedge (B \vee C) \dashv\vdash (A \wedge B) \vee (A \wedge C)$):

$$\begin{array}{lll} 1 & [1] & ((A \wedge X_1) \vee (B \wedge X_2) \vee X_3 \leftrightarrow C) \wedge ((C \wedge X_4) \vee (D \wedge X_5) \leftrightarrow E) & \text{A} \\ 1 & [2] & ((A \wedge X_1) \vee (B \wedge X_2) \vee X_3 \leftrightarrow C) \wedge (((A \wedge X_1) \vee (B \wedge X_2) \vee X_3) \wedge X_4) \vee (D \wedge X_5) \leftrightarrow E & \text{1, SUB\&} \\ 1 & [3] & ((A \wedge X_1) \vee (B \wedge X_2) \vee X_3 \leftrightarrow C) \wedge ((A \wedge X_1 \wedge X_4) \vee (B \wedge X_2 \wedge X_4) \vee (X_3 \wedge X_4) \vee (D \wedge X_5) \leftrightarrow E) & \text{2, DIS} \end{array}$$

The graphical notation illustrates this. The graph in figure 4.8 represents a possible expansion of graph (n). This expansion demonstrates that whatever factor Z is introduced into (n) above the C vertex, necessarily, is part of a minimally sufficient condition of E . This minimally sufficient condition containing Z is created by substituting C in CX_4 by whatever minimally sufficient condition of C Z happens to be part of.

In contrast to the structure-conserving extension \mathcal{F}_2 of \mathcal{F}_1 , extending \mathcal{F}_1 – interpreted along the lines of (II) – to \mathcal{F}_3 suspends the entanglement of C and E . In order to substantiate this, consider anew the minimal theory MT_o which represents

²⁰ Cf. section 3.2, p. 93.

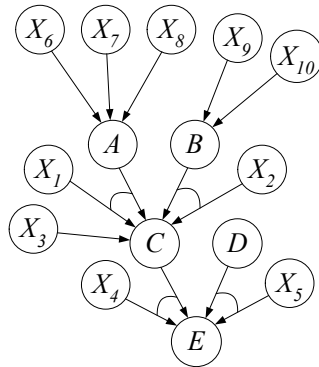


Fig. 4.8: A possible expansion of graph (n) in figure 4.7. Whatever is introduced into (n) above the C vertex is necessarily part of a minimally sufficient condition of E.

graph (o) of figure 4.7. Suppose, the factor frame of MT_o is extended and – among other things – additional factors are introduced into the set of minimally sufficient conditions of C, e.g. as follows:

$$(AX_1 \vee BX_2 \vee X_3 \Rightarrow C) \wedge (AX_4 \vee BX_5 \vee DX_6 \Rightarrow E) \quad (MT'_o)$$

From the mere form of MT'_o in combination with the transitivity of R it does *not* follow that whatever factor is part of a minimally sufficient condition of C must be part of a minimally sufficient condition of E as well. Or formally:

$$(AX_1 \vee BX_2 \vee X_3 \Rightarrow C) \wedge (AX_4 \vee BX_5 \vee DX_6 \Rightarrow E) \wedge \forall x,y,z (Rxy \wedge Ryz \rightarrow Rxz)$$

$\not\vdash AX_1 \mapsto E$	$\not\vdash X_{1/2}X_4 \mapsto E$	$\not\vdash X_3X_4 \mapsto E$
$\not\vdash BX_2 \mapsto E$	$\not\vdash X_{1/2}X_5 \mapsto E$	$\not\vdash X_3X_5 \mapsto E$
$\not\vdash X_3 \mapsto E$	$\not\vdash X_{1/2}X_6 \mapsto E$	$\not\vdash X_3X_6 \mapsto E.$

Contrary to the chain, an epiphenomenal structure does not necessitate newly introduced parts of minimally sufficient conditions of C – as X_1 or X_2 – to be contained

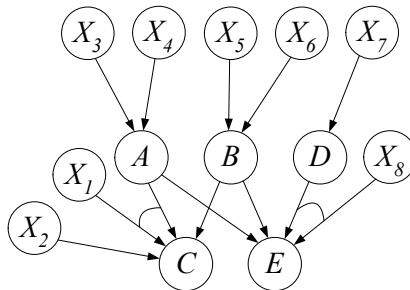


Fig. 4.9: A possible expansion of graph (o) in figure 4.7. Such an expansion suspends the entanglement of C and E. Neither X_1 nor X_2 are contained in minimally sufficient conditions of E.

in minimally sufficient conditions of E as well. The extension of graph (o) depicted in figure 4.9 illustrates this.

Even though every causal chain is *mt*-equivalent to an entangled epiphenomenon, these two structures differ with respect to expansions of their factor frames. While chains conserve entanglements across all factor frame extensions, entangled epiphenomena do not. Here lies the reason why we take interpretation (I) to describe a chain, whereas we hold the factors interpreted in terms of (II) to be underlain by an epiphenomenal structure. MT_o and graph (o) are thus extendable that there is at least one minimally sufficient condition for both C and E such that not all parts of that condition are contained in any minimally sufficient conditions of the other factor. If, however, no expansion of MT_o and graph (o) would result in a suspension of the entanglement of C and E , there would, even relative to interpretation (II), be no reason to model the causal structure generating R epiphenomenally.

The difference as regards the permanence of entanglements in (n) and (o) provides us, first, with a criterion that distinguishes chains and epiphenomena and, second, yields a rule that regulates the causal interpretation of entangled factors. If additional factors are integrated into a chain, entanglements are never resolved in principle. That is, if two factors Z_1 and Z_2 are entangled in a chain C , then Z_1 and Z_2 are entangled in *every* chain C' that results from expanding the factor frame of C by arbitrary factors. Therefore, we opt for the chain model in light of interpretation (I). If, however, two factors Z_1 and Z_2 are entangled in an epiphenomenal structure E , this entanglement is not preserved across all extensions of the factor frame of E . Ordinarily, the integration of additional factors into an epiphenomenon sooner or later resolves entanglements. That is why R is taken to be generated by an epiphenomenal structure when the factors in R are interpreted according to (II).

The retention of entanglements across arbitrary factor frame extensions is a characteristic feature of causal chains. In every chain there are at least two entangled factors whose entanglement subsists across all factor frame extensions. That is, such entanglements are necessary conditions for a respective coincidence frame to be the result of a causal chain. The same does not hold for epiphenomena. Even though certain factors may be entangled in epiphenomenal structures – as graph (o) demonstrates –, these structures do not necessitate such entanglements to be permanent. Epiphenomena are unambiguously identifiable only when there are minimally sufficient conditions for each effect, such that not all factors in the minimally sufficient conditions of the one effect are part of the minimally sufficient conditions of the other effect as well.

Therefore, even though modeling a coincidence frame as R in terms of a chain or an epiphenomenon is empirically equivalent and is thus equally warranted by R , it makes a crucial difference whether somebody causally analyzing that coincidence frame takes it to be generated by a chain or by an epiphenomenon. If he opts for the chain, he not only provides a causal model that is compatible with R , but moreover assesses that no factor newly introduced into this model will resolve the entanglement of C and E . Hence, taking a coincidence frame as R to be generated

by a chain is a *stronger claim* than modeling the causal structure behind R in terms of an epiphenomenon. The entanglement of C and E follows from the chain model, whereas that is not the case for the epiphenomenal model. In this sense, contrary to the latter, the chain model can be seen to *explain* the entanglement of C and E . The chain model exceeds the epiphenomenal model with respect to explanatory power. Entanglements can be represented, but not be explained by epiphenomenal models.

Notwithstanding the fact that both rival models are fully compatible with R , the choice between the chain and the epiphenomenon highly affects the course of the further analysis of a respective causal process. Trivially, extending the factor frame \mathcal{F}_1 either suspends the entanglement of C and E or it does not. If an extension of \mathcal{F}_1 leads to the discovery of a factor that is part of a minimally sufficient condition of C , but not part of one of E , the structure behind the behavior of the factors in \mathcal{F}_1 is proven to be epiphenomenal. Suppose, however, further investigation into the causal structure behind \mathcal{F}_1 does not reveal a previously unknown factor that suspends the entanglement of C and E . This scenario will be highly surprising to whomever has chosen the epiphenomenal model in the first place, for the epiphenomenal model does not provide any reason to expect the entanglement of C and E to subsist across factor frame extensions. In contrast, if the chain model is chosen to begin with, subsistence of entanglements will not come as a surprise at all. On the contrary, entanglements are a structural necessity of the chain model.

These considerations suggest the causal interpretability of entanglements that subsist across all factor frame extensions. That means, if two factors Z_1 and Z_2 are thus entangled that every part of a minimally sufficient condition of Z_1 is also part of a minimally sufficient condition of Z_2 and no extensions of the corresponding factor frame resolve this entanglement, Z_1 is causally relevant for Z_2 . According to this, entanglements that resist factor frame extensions are not only a necessary, but also a sufficient condition for a respective coincidence frame to be the result of a causal chain. The following criterion shall thus be taken to individuate causal chains:

Individuation of causal chains (ICC): The causal structure regulating the behavior of the factors in a frame \mathcal{F} is a chain iff there are at least two entangled factors in \mathcal{F} whose entanglement subsists across all factor frame extensions.

It needs to be emphasized, however, that the causal interpretability of permanent entanglements does not follow from coincidence data alone. It is not excluded on the basis of coincidence information alone that a coincidence frame with entanglements subsisting in light of arbitrary factor frame extensions is generated by an epiphenomenal structure. Though epiphenomena do not structurally necessitate entanglements to subsist across factor frame extensions, they do not necessitate entanglements to be suspended by such extensions either. Contrary to interpretation (II), it might – at least theoretically – happen that by whatever factor the frame of R is extended C and E remain entangled and, yet, still are parallel effects of the common causes A and B . Thus, while in ordinary epiphenomena entanglements

merely indicate that overly narrow factor frames have been analyzed, there might be anomalous epiphenomena such that no factor exists that resolves corresponding entanglements.

It is the explanatory power – and thus a non-empirical aspect – of the chain model that ultimately substantiates and justifies the modeling of coincidence frames, whose entanglements resist all factor frame extensions, in terms of chains. As long as entanglements are not suspended, there are no empirical grounds to model the causal structure behind a given coincidence frame in terms of an epiphenomenon. In this case, pending empirical grounds for an epiphenomenal modeling, explanatory power demands a modeling in terms of causal chains.

The persistence clause with respect to entanglements generates certain epistemic limitations as to the possibility of a conclusive identification of chains. Actual causal analyses always suffer from practical limitations as regards the complete expansion of investigated factor frames. Yet, before a factor frame has been fully expanded, (ICC) cannot irrefutably identify causal chains. According to the chain-problem there exists an *mt*-equivalent epiphenomenon to every chain, but not vice versa. The solution of the chain-problem induced by (ICC) only conclusively distinguishes between chains and epiphenomena relative to complete factor frames. Within fragmentary factor frames, (ICC) can only be resorted to as a negative criterion that strictly identifies coincidence frames that cannot be generated by a chain. Modeling the causal structure behind a given coincidence frame in terms of a chain is always fallible and needs to be relativized to the analyzed factor frame. Expansions of factor frames may constantly generate new empirical data that overthrows previous causal diagnoses.

Nonetheless, the above considerations show that even though the chain and the epiphenomenal model may empirically be equally warranted within fragmentary factor frames, the chain model represents a stronger claim with respect to factor frame extensions. Modeling the causal structure behind limited factor frames as in R in terms of chains, contrary to the epiphenomenal model, determines entanglements to be permanent. Moreover, as seen above, the chain model explains, rather than merely represents entanglements. Thus, both higher strength with respect to factor frame extensions and explanatory power on the side of the chain model suggest the following pragmatic inference rule when it comes to causally interpreting entanglements appearing within fragmentary factor frames: Entanglements are to be causally interpreted as long as they are not resolved by factor frame extensions. That is, if two factors Z_1 and Z_2 are thus entangled in \mathcal{F} that all factors contained in minimally sufficient conditions of Z_1 are part of minimally sufficient conditions of Z_2 , Z_1 is to be held causally relevant for Z_2 . This diagnosis, of course, is always to be relativized to the corresponding factor frame \mathcal{F} . If continued expansions of \mathcal{F} resolve the entanglement of Z_1 and Z_2 , a foregoing conclusion as to the causal relevance of Z_1 for Z_2 is proven to have been false.

According to (ICC) every causal chain comprises at least two entangled factors. Since the chain-problem only arises in light of coincidence frames that are compatible with chains, every instance of the chain-problem involves no less than two

entangled factors. The inference rule given above demands that this entanglement be causally interpreted as long as it is not resolved by factor frame extensions. Or put differently: Causal structures behind coincidence frames containing entanglements are to be modeled in terms of chains as long as a possibly underlying epiphenomenon has not been identified as an ordinary (non-entangled) epiphenomenon.

Causal interpretation of entanglements (CIE): If two factors Z_1 and Z_2 are thus entangled in a causally analyzed factor frame that all factors contained in minimally sufficient conditions of Z_1 are part of minimally sufficient conditions of Z_2 , only those minimally necessary conditions of Z_2 are to be interpreted causally that render Z_1 causally relevant for Z_2 . For short: Entanglements are to be interpreted causally.

In cases of the chain-problem, (CIE) asks for giving preference to the chain model. Hence, the causal structure generating R has to be modeled according to graph (n) – yet, only as long as no further factor has been discovered that is contained in a minimally sufficient condition of C , but not in one of E . As soon as factor frame extensions have resolved the entanglement of C and E , the relation between these two factors cannot be causally interpreted any longer. C and E will then be unambiguously identified as two parallel effects of a common cause.

(CIE) can be justified not only with reference to explanatory power or (ICC), but also by means of commonplace parsimony arguments. The reduction of a many- to a one-layer graph significantly increases the complexity of the respective causal structure. For instance, reconsider the reduction of the chain (l) to the epiphenomenon (m) in figure 4.6. This layer reduction involves a replacement of the single edge from D to G in (l) by a multitude of edges in (m) – one leaving from every factor causally relevant to D and entering G . (l) stipulates considerably less direct causal relevancies than (m). In order to predict the behavior of G , it suffices to supply information with respect to the behavior of D . Additionally taking account of A , B , C , or E is not called for. Causal knowledge serves the explanation and prediction of the behavior of corresponding factors. Such knowledge ought to be organized as simply as possible. By causally interpreting entanglements and thus preferring chains to entangled epiphenomena (CIE) does justice to these parsimony principles. In light of two different, yet *mt*-equivalent causal structures, (CIE) prefers the simpler to the more complex one.

4.5.2 Simple vs. Complex Minimal Theories (II)

The basic idea behind the regularity account of causation presented in this study consists in spelling out causal relevance with recourse to the persistent membership of factors in minimal theories. Causal relevance shall be defined in terms of ‘expansion resistant’ double-conditional dependencies among factors. Such dependencies hold iff of all logically possible coincidences within a given factor frame only a proper subset is empirically possible. Our definitional strategy will thus

only generate a well-defined notion of causal relevance, if *mt*-equivalent minimal theories, i.e. minimal theories compatible with identical coincidences, establish the same causal relevancies. It must be excluded that a factor is both assigned and not assigned causal relevance relative to its membership in causally differing, yet *mt*-equivalent minimal theories. However, as section 4.4 has shown, this condition is violated by MT_n and MT_o . The chain-problem hence induces an adaptation of the conceptual fundament of the regularity account introduced in the previous chapter.

Such a conceptual adaptation must refuse admittance into the set of causally interpretable minimal theories to one of two *mt*-equivalent, yet causally differing minimal theories. In light of (CIE), it is clear which of MT_n and MT_o is not to be causally interpreted: MT_o . Such as to maintain the tenet according to which causal relevance is to be defined by means of membership in minimal theories with persistent factor frames, the notion of a complex minimal theory introduced in section 3.6.5 (p. 136) shall be thus adapted that MT_o no longer satisfies the requirements imposed on complex minimal theories.

Simple and complex minimal theories (II):

1. A double-conditional with an antecedent consisting of a minimally necessary disjunction of minimally sufficient conditions and a consequent consisting of a single factor is a *simple* minimal theory of that consequent.
2. Every simple minimal theory is a minimal theory.
3. A conjunction of two minimal theories Φ and Ψ is a minimal theory iff
 - (a) at least one factor in Φ is part of Ψ ;
 - (b) Φ and Ψ do not have an identical consequent;
 - (c) for every j , $1 \leq j < n$, in a sequence of entangled factors Z_1, \dots, Z_n , $n \geq 2$: Z_j is contained in the antecedent of the simple minimal theory of Z_{j+1} ;
4. Minimal theories that are not simple are *complex*.

Condition (3c) excludes MT_o from the set of complex minimal theories, for C and E are entangled in R , yet neither of the two factors is contained in the antecedent of the simple minimal theory of the other. By refusing MT_o the status of a complex minimal theory, MT no longer both assigns and abstains from assigning causal relevance to C for E . C now is unambiguously causally relevant to E according to MT. In this manner, (ICC) and (CIE) are conceptually accounted for. Entanglements that remain unaltered by arbitrary factor frame extensions are to be interpreted causally. The conceptual ambiguities induced by the chain-problem are thus resolved.

4.6 The R-Solution?

This account of how causal chains and entangled epiphenomena are kept apart might be criticized for not conforming to how this distinction is drawn in everyday life. It might be argued that prevalent causal intuitions – for instance, as to modeling interpretation (I) in terms of a chain and interpretation (II) epiphenomenally – are not based on resistances of entanglements across factor frame extensions, but rather on differences between chains and epiphenomena with respect to certain (non-causal) *relational constraints* among causally linked events. There are several conceivable relational constraints that could be advanced as a means to solve the chain-problem. Let us consider them in turn.

4.6.1 Direction of Time

One suggestion to solve the chain-problem based on relational constraints could be to impose a chronological ordering onto the instances of the factors in a causal structure. If it is stipulated that causes always occur *before* their effects, a chain structure as graph (n) in figure 4.7 could be claimed to determine instances of *C* to occur *prior to* the instances of *E*. This, in turn, does not hold for the epiphenomenal structure depicted in graph (o). An entangled epiphenomenon as (o) is compatible with instances of *C* and *E* occurring *simultaneously*. Such as to illustrate this difference, graphs (n) and (o) are confronted with a timeline in figure 4.10.

Provided that causes are generally assumed to occur before their effects, the direction of time *prima facie* in fact seems to allow for an identification of chains and entangled epiphenomena. Furthermore, this criterion would even be empirically testable. Given a coincidence frame such as *R*, which could stem both from a chain and an entangled epiphenomenon, instances of *C* and *E* would simply have to be compared as to their chronological ordering. Finding *C* and *E* to be simultaneously instantiated would induce an inference to the epiphenomenal model. So far

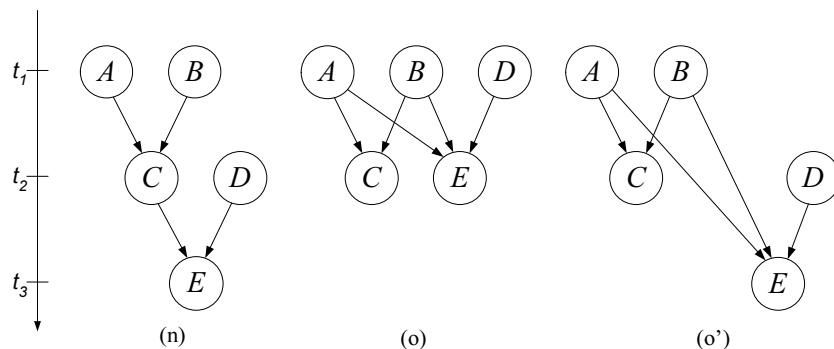


Fig. 4.10: The arrow on the left hand side represents the direction of time such that $t_1 < t_2 < t_3$. While (o) determines instances of *C* and *E* to occur simultaneously, (n) and (o') represent causal structures according to which *C* is instantiated before *E*.

so good, yet what if instances of *C* and *E* are not found to occur simultaneously? Would this constellation give preference to the chain model? As graph (o') in figure 4.10 shows, that is not the case. Epiphenomenal structures are *compatible* with simultaneous occurrences of their parallel effects, but they do *not determine* this simultaneity. Effects of a common cause might well be instantiated sequentially. Suppose, houses *r* and *s* in the power station example²¹ are connected to the power stations by wiring of different length or of different conductivity, such that electricity always reaches house *r* prior to house *s*. Nonetheless, of course, the causal structure behind this chronological specification of the power station example is to be modeled in terms of an epiphenomenon. That means, while (o) might be identifiable by means of chronological constraints, (n) and (o') cannot thus be kept apart, for they are not only *mt*-equivalent, but also chronologically equivalent. Every coincidence frame that could be the result of a chain might just as well be the product of a chronologically ordered entangled epiphenomenon of type (o'). Building the direction of time into a criterion that distinguishes between chains and entangled epiphenomena would merely allow for identifying those entangled epiphenomena that happen to be constituted by simultaneously occurring parallel effects. Such a criterion would, however, be of no help whatsoever when it comes to identifying causal chains. With recourse to the direction of time, the chain-problem could not be solved, but would just be specified.

Still, these considerations show that resorting to the direction of time and assuming the cause-effect relation to be sequentially ordered might yield a criterion that differentiates between chains and a subclass of entangled epiphenomena – the subclass with simultaneously occurring effects. This differentiation might be considered a sufficient reason to motivate such a recourse to the direction of time, even though the chain-problem is not solved in this vein. However, whether causes in fact always occur before their effects and thus whether a simultaneity of causes and effects is excluded on a priori grounds, is a question that is controversially discussed in the literature.²² Brand (1980), p. 142, for instance, mentions the following cases of what *prima facie* are examples of simultaneous causation:

- (63) The lead ball's going down causes the depressing of the cushion.²³
- (64) Jack's going down on the seesaw causes the seesaw's going down.
- (65) The increasing of the gravitational field causes the bending of the light beam.

In a similar vein, Huemer and Kovitz (2003) argue that many cases of simultaneous causation are backed and illustrated by the laws of classical physics. Clearly, if the assumption that causes always occur before their effects has to be given up, causal chains and entangled epiphenomena with simultaneous effects cannot be kept apart by means of chronological constraints any longer. For in this case, a coincidence

²¹ Cf. interpretation (II), p. 162 above.

²² Cf. Johnson (1963 (1924)), vol. III, p. xxv, Taylor (1966), pp. 35-39, Brand (1980), Hausman (1998), p. 44, Kistler (2001), p. 119, Huemer and Kovitz (2003).

²³ This often cited example derives from Kant (1956), *KrV*, A203.

frame as R , such that instances of C and E occur simultaneously, could both be the result of a chain and of an entangled epiphenomenon.

Yet, even if the assumption as to the sequential nature of all cause-effect relations should turn out to be somehow justifiable, distinguishing between chains and entangled epiphenomena with simultaneous effects by resorting to chronological constraints is hardly vindicable. It would amount to dismissing one of the central theoretical prospects the theory of causation proposed in this study substantiates. In section 3.6.2 we have seen that MT provides the means to analyze the direction of general causation on purely logical grounds and thus without recourse to external non-symmetries. This, in turn, paves the way for a causal analysis of the direction of time – all in line with the Reichenbachian programme. If the direction of time is now built into our criterion that distinguishes between chains and entangled epiphenomena with simultaneous effects, a causal analysis of the direction of time is rendered impossible.

4.6.2 Spatiotemporal Adjacency

The direction of time is not the only relational constraint that could be resorted to as a means to distinguish between chains and entangled epiphenomena. Spatiotemporal adjacency might be proposed as an alternative. This proposal would amount to fixing the interpretation of the relation R contained in the consequent of a minimal theory to "...is *spatiotemporally adjacent* to ...". As all thus far discussed interpretations of R , the notion of spatiotemporal adjacency is notoriously vague. It has to be understood differently when applied to microscopic and macroscopic processes or when applied to earthly events as opposed to cosmic ones. However, vagueness has not hindered us to introduce the fuzzy "...occurs in the same spatiotemporal frame as ..." as an interpretation of R in section 3.2. As a consequence of the vagueness of R , we have as yet settled for not letting R do any analytical work. The interpretation of R has therefore deliberately been left as open as possible. Causal analyses in general or the application of MT_d in particular have been relativized to the availability of a suitable interpretation of R given a respective factor frame. Abstaining from letting R do analytical work allowed for only imposing a negative condition on interpretations of R : R shall, relative to an analyzed factor frame, be thus interpreted that the factors in that frame are not prevented from causal interaction merely due to a lack of R -relatedness of their instances.

This noncommittal stance with respect to R could not be upheld any longer, if an R interpreted in terms of spatiotemporal adjacency were now implemented as a criterion that distinguishes chains and entangled epiphenomena. Such a criterion would call for imposing positive conditions on interpretations of R , for the criterion would only be applicable if for any two events it were clear whether they are related in terms of R or not, i.e. whether they are spatiotemporally adjacent or not. Thus, when confronted with a coincidence frame as R in section 4.4 the instances of

the involved factors would have to be tested for spatiotemporal adjacency. Whenever these instances are not adjacent, they could not be directly causally related by definition.

In light of the fact that spatiotemporal intervals in which causally dependent factors are instantiated highly vary with the causal process at hand, having to positively fix the interpretations of R to specific intervals is a serious theoretical drawback. Since R is a constitutive element of our analysis of causation, its interpretations cannot be allowed to presuppose causal notions. Hence, an interpretation rule for R calling for interpretations of R that depend on the *causal* process under investigation is unfit. Alternatives, however, are not available in the literature nor do they lay at hand. Nonetheless, in order to see whether the chain-problem could be solved along these lines, let us assume that for any given factor frame it is sufficiently clear and non-causally determinable when factors are instantiated adjacently and when they are not.

Against this background, the following solution to the chain-problem as posed by the coincidence frame R might be proposed:

- (i) The causal structure generating R is a chain iff $A \vee B$ is causally relevant to C and $C \vee D$ to E and instances of $A \vee B$ and C , on the one hand, and of $C \vee D$ and E , on the other, are spatiotemporally adjacent.
- (ii) The causal structure generating R is an entangled epiphenomenon iff $A \vee B$ is causally relevant to C and $A \vee B \vee D$ to E and instances of $A \vee B$ and C , on the one hand, and of $A \vee B \vee D$ and E , on the other, are spatiotemporally adjacent.

Now let us consider the two interpretations that have been given to the factors in R above. What could “spatiotemporally adjacent” mean relative to the factors involved in interpretation (I), i.e. what does it mean for the turning of the starter key to be adjacent to the spark plug and for the running car engine to be adjacent to the rolling of the wheels? Let us assume that the turning of the starter key occurs within 2 seconds and 50 cm of the spark plug firing and that the running of the car engine is also located within 2 seconds and 50 cm of the rolling of the wheels. Spatiotemporal adjacency shall thus be taken to mean “. . . occurs within 2 seconds and 50 cm of . . .”. This certainly is an implementable interpretation of R. According to (i) and (ii), the causal structure behind R, interpreted in terms of (I), is a chain if the turning of the starter key is in fact located within 2 seconds and 50 cm of the firing spark plug, yet is not thus adjacent to the rolling wheels. In an ordinary car this normally will be the case. Is this the reason why we intuitively model the structure behind R – interpreted in terms of (I) – as a causal chain and not as an epiphenomenon?

This question could be positively answered if we were in fact ready to model the structure behind the car example in terms of an epiphenomenon, given that the involved factors were not instantiated thus that their instances are spatiotemporally

related according to what is stipulated in (i), but rather according to what is stipulated in (ii). Hence, let us assume that some outlandish engineer designed a car whose starter lock is located on the hubcap of the right rear wheel. Yet apart from this, the remarkable car under consideration resembles ordinary cars in all detail, i.e. the starter lock on the hubcap is wired to the spark plug and the engine is mechanically connected to the axis. That means, the turning of this car's starter key does not happen within 50 cm of the firing spark plug, but within 50 cm of the rolling wheels. This extraordinary car thus constitutes an instance of a constellation as described in (ii). Is the turning of the starter key now directly causally relevant to the car movement, i.e. without intermediation by the engine? Would we be ready to model the causal structure behind the movement of this car in terms of an entangled epiphenomenon? Certainly not.

It might be argued that in view of such a car the notion of spatiotemporal adjacency needs to be adapted. However, what spatiotemporal interval could be chosen such that the turning of the starter key would be adjacent to the firing spark plug but not to the rolling wheels? Since the starter lock is spatially farther away from the spark plug than from the wheels, there obviously is no such interval that would render this car's movement in line with (i) and thus with the spatiotemporal constraints imposed for a causal chain. Nonetheless, turnings of starter keys do not directly cause car movements – how ever adjacent starter locks and rolling wheels might be. It is not the spatiotemporal relationship between the involved factors that determines the causal process behind moving cars to be structured in terms of chains.

The situation is similar in case of entangled epiphenomena. Consider anew the power supply example, i.e. interpretation (II) of the factors in R . What would be an adequate interval for spatiotemporal adjacency relative to the factors involved in this example? Let us assume that operating power stations and power supplied houses are located within 3 seconds and 5 km of each other in the city under consideration. Now suppose that houses r and s are situated at different ends of this city, say 10 km apart, whereas they both are located within the 5 km distance from the two power stations. This clearly is a constellation as described in (ii). The power supply in house r cannot be directly causally relevant to the power supply in house s , thus spatiotemporal constraints impose to model this process epiphenomenally.

The objection against this analysis is at hand. Supposing that houses r and s are neighboring does not change the causal structuring of the underlying process at all. Given that the two houses are neighboring does not turn the epiphenomenon into a chain. Spatiotemporal adjacency of the instances of the involved factors is simply irrelevant as to whether investigated causal processes are chainlike or epiphenomenally structured. The graphs in figure 4.11 illustrate this. Besides causal dependencies, these graphs represent the spatiotemporal distances between the instances of the factors A to E . The measure labelled "R" marks the distance within which instances of factors can be seen as spatiotemporally adjacent. That means e.g. for graph (n) that instances of A and B are adjacent to events of type C and that instances of C and D are adjacent to events of type E . This constellation conforms to

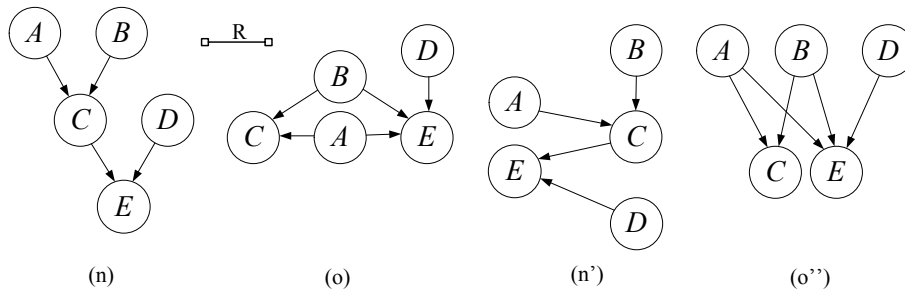


Fig. 4.11: Causal graphs that not only represent causal dependencies, but also mirror the spatiotemporal distance between instances of respective factors. All factors that are not located farther apart than the distance labeled by “R” have to be seen as spatiotemporally adjacent.

(i) and thus induces a modeling in terms of a chain. In contrast, the causal and spatiotemporal relationships depicted in graph (o) comply with (ii) and hence with an epiphenomenon. That means, (n) and (o) in figure 4.11 coincide with (n) and (o) in figure 4.7 – thence the identical labels –, except for the additional spatiotemporal information included in 4.11. However, as graphs (n’) and (o’’) indicate, changing spatiotemporal relationships do not alter causal dependencies. By bringing the starter lock and the rolling wheels closer together, as symbolized in (n’), the causal structure does not change from chain to epiphenomenon. Correspondingly, by investigating the power supply in two neighboring houses, an epiphenomenon is not transformed into a causal chain.

This rejection of (i) and (ii) as sustainable criteria that identify chains and epiphenomena might be criticized for being too coarse-grained. Spatiotemporal adjacency, it might be argued, can only be understood in terms of spatiotemporal *contiguity*, meaning that events have to be in direct spatiotemporal contact in order to causally interact. This yields the following accentuation of (i) and (ii):

- (i’) The causal structure generating R is a chain iff $A \vee B$ is causally relevant to C and $C \vee D$ to E and instances of $A \vee B$ and C , on the one hand, and of $C \vee D$ and E , on the other, are spatiotemporally contiguous.
- (ii’) The causal structure generating R is an entangled epiphenomenon iff $A \vee B$ is causally relevant to C and $A \vee B \vee D$ to E and instances of $A \vee B$ and C , on the one hand, and of $A \vee B \vee D$ and E , on the other, are spatiotemporally contiguous.

(i’) and (ii’), of course, are neither satisfied for the instances of the factors involved in the car example nor for the ones involved in the power station example, which shows – thus the argument continues – that an overly coarse-grained factor frame is chosen for these examples. In order to obtain appropriate factor frames, additional factors, which are located on a causal path in between the coarse-grained factors A and E , have to be taken into account. In this sense, for instance, a factor representing suitable wiring connecting the starter lock to the spark plug would have to be

introduced in case of the car example, whereas adequate supply lines are required in case of the power station example. Given such supply lines from the power stations to the two houses, the criterion of spatiotemporal contiguity can be held to be satisfied, for now the power production in both plants is contiguous to the supply lines which, in turn, are contiguous to the power supplied houses. In this respect, there is a causal path composed of contiguously instantiated factors that leads from the power production to the power suppliance of the two houses. Hence, (i') and (ii') are to be specified as follows:

- (i'') The causal structure generating R is a chain iff $A \vee B$ is causally relevant to C and $C \vee D$ to E and instances of $A \vee B$ and C , on the one hand, and of $C \vee D$ and E , on the other, are spatiotemporally contiguous or are connected by a causal sequence of spatiotemporally contiguously instantiated factors.
- (ii'') The causal structure generating R is an entangled epiphenomenon iff $A \vee B$ is causally relevant to C and $A \vee B \vee D$ to E and instances of $A \vee B$ and C , on the one hand, and of $A \vee B \vee D$ and E , on the other, are spatiotemporally contiguous or are connected by a causal sequence of spatiotemporally contiguously instantiated factors.

Contrary to the argument put forward against (ii), by assuming the houses to be adjacent nothing changes with respect to the satisfaction of (ii''). As long as there is no supply line between neighboring houses, these power supplied houses do not violate (ii'') and thus the causal process behind their power suppliance remains amenable to an epiphenomenal modeling. Indeed, if there were a supply line between the houses, it might in fact make a crucial difference as to whether the corresponding process is chainlike or epiphenomenally structured. Given a supply line between the houses, it might well be the case that the power supply in one house is now directly causally relevant to the power supply in the other house.

This *prima facie* seems to be strong support for the thesis that chains and epiphenomena are in fact identified by criteria along the lines of (i'') and (ii''). Let us therefore consider how (i'') and (ii'') perform when applied to the car example. As mentioned above, it is beyond doubt that turnings of starter keys are not contiguous to firing spark plugs, which, in turn, are not contiguous to rolling wheels. As soon as this coarse-grained factor frame, however, is extended by the wiring that transmits the electrical impulse from the starter lock to the spark plug and by the mechanism that confers the kinetic energy from the engine to the axis and the wheels, (i'') becomes satisfied. This factor frame extension yields a sequence of spatiotemporally contiguously instantiated factors from A or B to E , such that each event of the sequence causes its successor.

This shows that distinguishing between chains and entangled epiphenomena with recourse to criteria as (i'') and (ii'') would imperatively induce analyzing causal structures on a fairly fine-grained level of specification – additionally to providing some as yet unavailable non-causal interpretation rule for R . Among factors whose instances are not contiguous, (indirect) causal dependencies could not be

investigated to begin with. The level of specification of causal analyses would not be open and adjustable to the interests and requirements of a given causal investigation, but would a priori be fixed to a level thus fine-grained that factor frames contain sequences of spatiotemporally contiguously instantiated factors. Furthermore, it is far from clear whether such contiguous sequences of events can be found in case of all causal processes. Consider, for instance, the elliptical motion of the planets caused by the gravitational forces between the planets and the sun or the budget deficit of a firm which causes the firm to lay off a certain number of its employees. In what sense can the instances of these causally dependent factors be seen as connected by sequences of contiguous events? Spatiotemporal contiguity of events thus is not a necessary condition for corresponding factors to be causally related. There are causal structures whose factors do not form sequences such that neighbors are contiguously instantiated.

Moreover, criteria along the lines of (i'') and (ii'') neither are sufficient conditions for factors as *A*, *B*, *C*, *D*, and *E* to be causally structured in terms of chains and epiphenomena, respectively. This is illustrated by the remarkable car with its starter lock installed on its hubcap. On this car the turning of the starter key and the rolling wheels are spatiotemporally contiguous, and still the former factor is not directly causally relevant to the latter. If the distinction between chains and entangled epiphenomena really hinges on some form of spatiotemporal contiguity, it must be a very special kind of contiguity that accounts for this distinction. It must be a form of contiguity such that 'causal influence' is transmitted from one event to its contiguous successor. In case of the car with the starter lock on its hubcap such transfer of 'causal influence' proceeds through the wiring and by mediation of the spark plug and not directly from the lock to the wheels. That is the reason why the functioning of this extraordinary car has to be modeled in terms of a chain and not of an epiphenomenon. Yet, a notion of spatiotemporal adjacency or contiguity that involves 'causal influence' or any other causal predicate, obviously, is of no use when it comes to a conceptual analysis of the notion of direct causal relevance – and nothing less is at stake here.

These considerations demonstrate that not only interpretations of spatiotemporal adjacency call for recourse to causal notions, but also interpretations of spatiotemporal contiguity are closely interwoven with causal conceptions. Only instances of factors are contiguous in the sense needed for a solution of the chain-problem, which are neighbors on a causal path – transference of 'causal influence' being a precondition for contiguity among factors. Hence, there is no way around interpreting spatiotemporal contiguity, i.e. R as interpreted in (i'') and (ii''), with recourse to causal notions. This, in turn, forecloses resorting to R when it comes to analyzing (direct) causal relevance. The difference between chains and epiphenomena is to be accounted for by non-causal means. Such means are not provided by fixing the interpretation of R to "...is spatiotemporally adjacent/contiguous to ...", for these relations – in the sense needed for solving the chain-problem – are merely seemingly non-causal. Specifying conditions that allow for determining concrete spatiotemporal intervals for "...is spatiotemporally adjacent/contiguous

to ...” relative to a given factor frame inevitably involves clarity about the causal process under investigation, which, again, this contiguity relation is just designed to provide.

4.6.3 Energy Transfer

Authors as Aronson (1971) and Fair (1979) have proposed to resort to *energy* transfer as a criterion that identifies causal processes. The so-called *transference theory* of causation, which in modern variants is propagated e.g. by Dowe (2000), essentially claims that two events are causally related iff there is a transfer of energy or momentum from the one event to the other.²⁴ Such as to avoid definitional circles as regards a possible R-solution of the chain-problem, it might thus be proposed to interpret R in terms of “there is an energy transfer from ... to ...”. This would yield the following criteria determining the causal modeling of the factors in R:

- (i’’) The causal structure generating R is a chain iff $A \vee B$ is causally relevant to C and $C \vee D$ to E and there is energy transfer from the instances of $A \vee B$ to the ones of C and from the instances of $C \vee D$ to the ones of E .
- (ii’’) The causal structure generating R is an entangled epiphenomenon iff $A \vee B$ is causally relevant to C and $A \vee B \vee D$ to E and there is energy transfer from the instances of $A \vee B$ to the ones of C and from the instances of $A \vee B \vee D$ to the ones of E , but not from the instances of C to the instances of E .

The restriction as to the absence of energy transfer from C to E is added in (ii’’) because, other than the interpretations of R encountered in the previous section, energy transfer could be taken to be a transitive relation.²⁵ In this case, a chain in the sense of (i’’) would automatically satisfy a variant of (ii’’) without that restriction as well, which, accordingly, could not serve as an identifier of epiphenomena.

By means of (i’’) and (ii’’) interpretations (I) and (II) of the factors in R are modeled in accordance with common causal judgments. There is a transfer of energy from the operating engine to the rolling wheels, which, therefore, is a chain structure. In contrast, no energy is transferred from house r to house s, which renders the power station example in line with (ii’’) and thus induces an epiphenomenal modeling. This successful discrimination between chains and entangled epiphenomena is unaffected by any of the modifications of the examples discussed in the previous section. Hence, even the causal structuring of the car, whose starter lock is located on a hubcap, is identified to be a chain by (i’’), because energy is

²⁴ Energy and momentum are not the only quantities put forward as being the transferred entities in causal processes. Depending on the process under investigation, different quantities are being transferred. In order to remain non-committal as regards transferred quantities, modern transference theoretic accounts speak of *conserved quantities* as being the objects of transference (cf. Dowe (2000) or Kistler (2001)). For simplicity’s sake and such as to not having to define the notion of a conserved quantity, I shall just make reference to energy transfer as a wild card for all kinds of transferred quantities.

²⁵ The transitivity of energy transfer is not undisputed (cf. Ehring (1997), ch. 3).

transferred from the turning of the lock to the spark plug and from the operating engine to the wheels.

No doubt, the transference theory of causation captures a widespread intuition as regards the causal relation. Causes are commonly taken to ‘do’ something to their effects, and, of course, many causal processes – mainly the ones belonging to the field of physical sciences – in fact feature energy transfers from causes to effects. Does an interpretation of R in terms energy transfer thus solve the chain-problem in a way that is more in line with how chains and epiphenomena are kept apart in everyday life than a solution based on (ICC)?

Transference theories of causation have been criticized in twofold ways. First, it has been pointed out that by far not all causal processes in fact involve energy transfer, and second, transference theories have been rejected for being circular.²⁶ Consider the following scenario: Fennella turns off the light in the kitchen and shortly thereafter Hogan stumbles over an empty beer bottle lying on the kitchen floor. Fennella’s turning off of the light is part of the cause of Hogan’s stumbling, yet there is no energy transfer from the turning of the light switch to Hogan’s mishap. On the contrary, it can be argued that turning the light switch *intercepts* an ongoing energy transfer. According to a criterion as (ii’”), however, Fennella’s action and Hogan’s accident can, at best, be effects of a common cause. This is not an awkward singular scenario. There is a host of causal processes not featuring energy transfers.²⁷ Infection by a flu virus is a cause of an influenza without the virus transferring energy to the body. Or drunken driving causes accidents, yet there is no energy transfer from the alcohol consumption to the crash. Moreover, the whole field of social or historical processes is not amenable to a causal interpretation relative to a transference theoretic account. What kind of energy transfer took place from the causes of the French Revolution to the fall of the Bastille?

Furthermore, it is doubtful whether the transference theoretic analysis of causation does not in fact presuppose causal notions. For what are transference processes, irrespective of what entity is being transferred, other than causal processes? In order to make sense of the notion of a transference process, clarity on what a causal process is needs to be provided. Transference processes are only identifiable given an analysis of causation, and the latter is just what the transference theory intends to supply. However, let us for the moment ignore these conceptual hurdles in order to see whether, given that clarity on the notion of a transfer process could somehow be presupposed, the chain-problem could at least be solved for physical processes that involve energy transfer. According to such a solution, whether the structure behind coincidence list R is a chain or an epiphenomenon is to be determined by checking if there is energy or some other conserved quantity being transferred from the instances of *C* to the instances of *E*. That, in turn, requires that some kind of measurement device is installed on the path from *C* to *E*. If this device detects energy transfer, the structure generating R is a chain, otherwise

²⁶ Cf. Dowe (2000), Kistler (2001).

²⁷ Cf. Schaffer (2000).

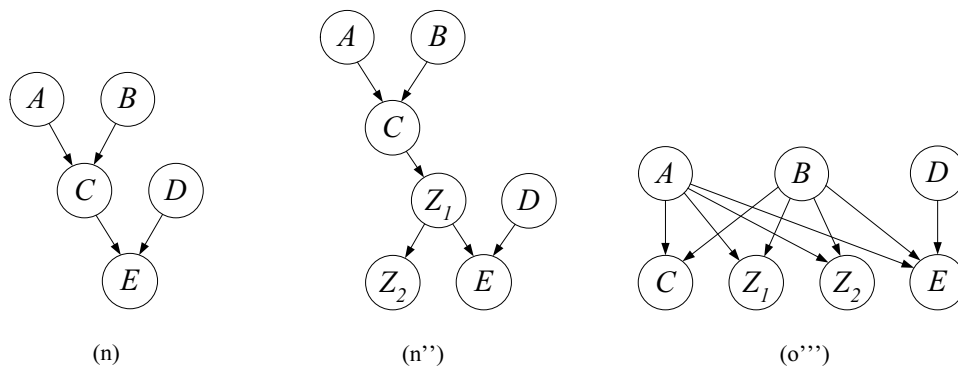


Fig. 4.12: Expanding the chain model (n) by introducing factors on the path between C and E does not solve the chain problem. To (n'') there again exists an *mt*-equivalent epiphenomenon: (o''').

it turns out to be an epiphenomenon. Installing such a device amounts to an expansion of the causal model depicted in graph (n). In graph (n'') of figure 4.12 this expansion is illustrated by the introduction of factors Z_1 and Z_2 , where Z_1 represents some conserved quantity being transferred and Z_2 stands for a corresponding measurement device.. Clearly, (n'') is not *mt*-equivalent to the entangled epiphenomenon (o). Causal structures with different factor frames, trivially, do not generate identical coincidence frames. Yet, as (n'') is a causal chain as well, there again exists an *mt*-equivalent epiphenomenon, *viz.* (o'''). Thus, while a transfer criterion may distinguish between (n'') and (o), it does not distinguish between (n'') and (o'''). Now, of course further measurement devices might be installed on $\langle C, Z_1 \rangle$ and $\langle Z_1, E \rangle$, respectively. Yet, for all such expansions of (n'') there will exist *mt*-equivalent epiphenomena. That means, even if the notion of a transfer process could somehow be non-causally clarified, the chain-problem could not be solved by drawing on that notion. Rather than solving the chain-problem a transfer criterion along the lines of (i''') and (ii''') would, at best, call for further and further expansions of investigated factor frames without ever positively identifying a causal chain.

4.6.4 Non-transitive R

The above considerations have shown that fixing the interpretation of R to specific relations as spatiotemporal contiguity or energy transfer does not adequately solve the chain-problem, the main reason for this inadequacy being the implicitly causal nature of these relations. Conceding that interpretations of R need to remain variable depending on the factor frame under investigation, it might still be argued that possible interpretations of R can be narrowed down in a way that solves the chain-problem. One possible restriction as to admissible interpretations might exclude

all transitive interpretations of R. Hence, what about fixing R to be a non-transitive relation?²⁸

This proposal might be backed by contending that the chain-problem essentially evolves from the transitivity of the two central dependency relationships implemented in MT: "... is part of a minimally sufficient condition of ..." and "... is part of a minimally necessary condition of ...". Prescinding from relational constraints, the following are theorems:

$$\begin{aligned} \forall x((Ax \vee Bx) \rightarrow \exists yCy) \wedge \forall x((Cx \vee Dx) \rightarrow \exists yEy) \vdash \\ \forall x((Ax \vee Bx \vee Dx) \rightarrow \exists yEy) \end{aligned} \quad (4.1)$$

$$\begin{aligned} \forall x(Cx \rightarrow \exists y(Ay \vee By)) \wedge \forall x(Ex \rightarrow \exists y(Cy \vee Dy)) \vdash \\ \forall x(Ex \rightarrow \exists y(Ay \vee By \vee Dy)) \end{aligned} \quad (4.2)$$

If A and B are each minimally sufficient for C and C and D are each minimally sufficient for E , A and B are minimally sufficient for E as well. Analogously, if $A \vee B$ is minimally necessary for C and $C \vee D$ is minimally necessary for E , $A \vee B \vee D$ is minimally necessary for E . This transitivity, it might be argued, lies at the heart of the chain-problem. For (4.1) and (4.2) basically state that from the dependencies holding among the factors in a chain the dependencies holding among the factors in an entangled epiphenomenon can be derived. (4.1) and (4.2), however, do not conform to the restricted notions of minimal sufficiency and necessity introduced in sections 3.3 and 3.4. One crucial additional requirement imposed on causally interpretable sufficient and necessary conditions demands that their instances be properly related in space and time. This is an important restriction, because from A and B being minimally sufficient for C and properly related to C and C and D being minimally sufficient for E and properly related to E it does *not follow* that A and B are minimally sufficient for E and properly related to the latter. A and B , though minimally sufficient for E in the customary logical sense, are not necessarily instantiated by events that are properly spatiotemporally related to the instances of E . Nonetheless, if proper spatiotemporal relatedness is taken to be transitive, the derivative relationships as expressed in (4.1) and (4.2) reappear:

$$\begin{aligned} \forall x((Ax \vee Bx) \rightarrow \exists y(Cy \wedge Rxy)) \wedge \forall x((Cx \vee Dx) \rightarrow \exists y(Ey \wedge Rxy)) \wedge \\ \forall x \forall y \forall z (Rxy \wedge Ryz \rightarrow Rxz) \vdash \forall x((Ax \vee Bx \vee Dx) \rightarrow \exists y(Ey \wedge Rxy)) \end{aligned} \quad (4.3)$$

$$\begin{aligned} \forall x(Cx \rightarrow \exists y((Ay \vee By) \wedge Rxy)) \wedge \forall x(Ex \rightarrow \exists y((Cy \vee Dy) \wedge Rxy)) \wedge \\ \forall x \forall y \forall z (Rxy \wedge Ryz \rightarrow Rxz) \vdash \forall x(Ex \rightarrow \exists y((Ay \vee By \vee Dy) \wedge Rxy)) \end{aligned} \quad (4.4)$$

(4.3) and (4.4) demonstrate that, given a transitive R, the dependencies among the factors in a causal chain imply the dependencies among the factors in an entangled epiphenomenon. The crucial premiss in both (4.3) and (4.4) is the transitivity

²⁸ For details on the notion of non-transitivity implemented here cf. Lemmon (1978 (1965)), p. 183.

of R . If, however, R is taken to be non-transitive, chains do not imply entangled epiphenomena in the sense expressed by (4.3) and (4.4). Therefore, narrowing interpretations of R down to non-transitive relations *prima facie* seems to be a promising way around the chain-problem. What about the merits of this proposal?

As in case of the fixed interpretations of R considered in the previous sections, excluding all transitive interpretations amounts to letting R do the decisive work when it comes to identifying chains and entangled epiphenomena. Contrary to the previously proposed R -solutions of the chain-problem, the current proposal, however, does not fix R to a specific interpretation. It therefore does not seem to be affected by the difficulties that arise from determining a specific spatiotemporal interval, within which events can be seen as R -related, or from accounting for transfer processes non-causally. Yet, this apparent difference between the current and the foregoing R -solutions disappears upon a closer look. A criterion that distinguishes between chains and epiphenomena needs to be applicable in the course of causal reasoning. When confronted with the behavior of the factors in a concrete frame, one has to be able to test for R -relatedness of investigated events. Only thus, according to the proposal under consideration, is it possible to keep chains and epiphenomena apart. Clearly, knowing that R is non-transitive is not sufficient for evaluating whether concrete events in fact are related in terms of R or not. Identity criteria for chains and entangled epiphenomena relying on relational constraints among instances of causally dependent factors are only operative provided that for any pair of events $\langle e_1, e_2 \rangle$ it is determinable whether Re_1e_2 or $\neg Re_1e_2$. To this end, an established non-transitivity of R evidently is insufficient. Any solution of the chain-problem that essentially hinges on relational constraints imposed on causally related events cannot evade providing operative interpretations of these constraints. This, in turn, amounts to nothing less than to administering sufficient and necessary conditions for the R -relatedness of pairs of events, which imposes the exact same difficulties on this alleged R -solution of the chain-problem as were encountered upon considering the previous proposals.

Furthermore, merely excluding transitive interpretations of R cannot amount to distinguishing between chains and entangled epiphenomena, for, as we have seen in section 4.6.2, there are several non-transitive interpretations of R that are no good in solving the chain-problem. Non-transitive relations as spatiotemporal adjacency or contiguity only serve the identification of chains and entangled epiphenomena if they are implicitly causally interpreted, which forecloses their applicability within analyses of causal relevance. Hence, more would need to be said about R than merely excluding transitive interpretations. Yet, as we have seen, fixing interpretations of R to specific relations in a way that is suitable to solve the chain-problem is a highly difficult task. Interpretations of R that could function as identity criteria for chains and entangled epiphenomena depend on the causal process under investigation, which, according to the proposal at hand, is supposed to be modeled with recourse to just that relation.

Finally, fixing interpretations of R to non-transitive relations conflicts with the most popular interpretation found in the literature: chronological order, i.e.

“...occurs before...”. Of course, if the chain-problem were solvable by fixing R to some non-transitive relation, the uncommon nature of this proposal would not tip the scales against it. Nonetheless, this shows that excluding all transitive interpretations of R cannot even draw on a certain intuitive plausibility.

Analyzing causation in terms conceptions such as regularities or probabilistic and counterfactual dependencies requires to impose non-causal relational constraints upon causally related events. Regularities or probabilistic and counterfactual dependencies may subsist among any events, irrespective of their spatiotemporal distance. Yet, only a subset of all these dependencies is causally interpretable. The relational constraints are to be thus defined that they select that subset. The nature of these constraints, which in the present contexts are formally represented by relation R, is highly sensitive to the process under investigation. R receives completely different interpretations in the domain of micro processes as compared to causal interactions on macro level. Due to the impossibility to fix interpretations of R to specific relations, only a negative condition has been imposed on R so far: R is to be interpreted thus that causal dependencies among factors in analyzed frames are not excluded merely due to a lack of R-relatedness of respective instances. As a consequence of this vagueness, no analytical function has been assignable to R – neither when it came to determining the direction of causation nor, now, in light of the chain-problem. Even though a solution of the chain-problem drawing on relational constraints may seem intuitively appealing, it would amount to assigning a decisive analytical function to R. R would become the core of a criterion distinguishing between chains and entangled epiphenomena. This, in turn, would presuppose necessary and sufficient conditions for interpretations of R, which this section has not succeeded in providing.

4.6.5 *Summing Up*

This chapter has shown that the most common causal structure – chains – imposes serious problems on currently available theories of causation. The chain-problem as encountered by a regularity account based on MT essentially consists in the *mt*-equivalence of chains and entangled epiphenomena. The solution to the problem drawing on (ICC) primarily rests on making suitable use of factor frame extensions, which constitute an analytical means already resorted to in MT_d and MT_i . Entanglements are to be causally interpreted iff they subsist across all factor frame extensions. (ICC) has at least three crucial advantages over all conceivable R-solutions. First, (ICC) neither explicitly nor implicitly presupposes any causal notions. It thus guarantees for an identification of chains and entangled epiphenomena in a non-circular way. Second, it dispenses with having to fix interpretations of R to specific relations. Contrary to all possible R-solutions, (ICC) does not assign any analytical function to R, whose vagueness hence does not have to be eliminated. And third, (ICC) does not introduce any analytical means into MT that have not been resorted to before. It is thus fully in line with the theoretical framework of MT.

4.7 Causal Interpretation of Minimally Necessary Conditions

Section 4.4 has pointed out that the chain-problem can also be seen as a minimalization problem with respect to necessary conditions. Factors A , B , and C of the exemplary coincidence frame R are each minimally sufficient for E . A and B , however, are never instantiated without a concomitant instantiation of C . Therefore, $C \vee D$ and $A \vee B \vee D$ both constitute minimally necessary conditions of E . This example reveals that effects sometimes have several different minimally necessary conditions, not all of which are amenable to a causal interpretation, for they would be contradictory in causal respects. Minimalizing necessary conditions thus is not generally unambiguous. Though there is at least one minimally necessary condition for each effect, there is *not* necessarily *exactly one* such condition for each effect. The chain-problem demonstrates that for every effect located at the tail of a causal chain there are at least two *mt*-equivalent, yet causally inequivalent double-conditionals.

Tails of causal chains are not the only effects whose necessary conditions are not unambiguously minimizable. There also are effects contained in one-layer²⁹ structures whose behavior can be represented by several *mt*-equivalent double-conditionals. Following Quine (1959), Kim (1993) has constructed a case of this sort.³⁰ Suppose, there are the following four minimally sufficient conditions for an effect W : $A\bar{B}$, $\bar{A}B$, $A\bar{C}$ and $B\bar{C}$. The disjunctive concatenation of these conditions is not minimally necessary, i.e.

$$A\bar{B} \vee \bar{A}B \vee A\bar{C} \vee B\bar{C} \Rightarrow W \quad (4.5)$$

is no minimal theory of W . The antecedent of (4.5) contains redundant disjuncts. The disjunction resulting from eliminating either $A\bar{C}$ or $B\bar{C}$, yet not both, remains necessary for W . Minimalizing the antecedent of (4.5) thus yields the following minimal theories of W :

$$A\bar{B} \vee \bar{A}B \vee A\bar{C} \Rightarrow W \quad (MT_p)$$

$$A\bar{B} \vee \bar{A}B \vee B\bar{C} \Rightarrow W \quad (MT_q)$$

MT_p and MT_q are *mt*-equivalent, yet, if causally interpreted, identify different complex causes of W .

This difficulty *prima facie* seems related to the chain-problem. However, the impossibility to unambiguously minimalize the antecedent of (4.5) results from *logical* dependencies among its disjuncts, whereas in case of chains the analogous minimalization problem is generated by *causal* dependencies among minimally sufficient conditions. The following conditional dependencies hold among the disjuncts of (4.5):

$$A\bar{C} \rightarrow B\bar{C} \vee A\bar{B} \quad (4.6)$$

$$B\bar{C} \rightarrow \bar{A}B \vee A\bar{C} \quad (4.7)$$

²⁹ For details on the notion of a layer of a causal structure cf. p. 137.

³⁰ Cf. Kim (1993), p. 67.

For every instance of $A\bar{C}$ there is an instance of $B\bar{C} \vee A\bar{B}$ and for every instance of $B\bar{C}$ there is an instance of $\bar{A}B \vee A\bar{C}$. That means, whenever $A\bar{C}$ or $B\bar{C}$ are instantiated, one of the other disjuncts of the antecedent of (4.5) is instantiated as well.

Do these ambiguities with respect to the minimalization of (4.5), notwithstanding their purely logical origin, constitute a problem for our theory of causation similar to the chain-problem? On the face of it, this seems to be the case. For a causal interpretation of the *mt*-equivalent expressions MT_p and MT_q would yield different causal structures, which coincides with the ambiguities as generated by MT_n and MT_o . Double-conditionals regulating the behavior of the factors in a causal chain and in an entangled epiphenomenon are *mt*-equivalent as well, and, if causally interpreted, exhibit different causal structures. However, the logical origin of the ambiguous minimalization of (4.5) renders this ambiguity unproblematic for our analysis of causal relevance. For logical reasons, $A\bar{C}$ or $B\bar{C}$ can never be instantiated independently of other disjuncts in the antecedent of (4.5). According to the Principle of Relevance (PR)³¹, however, without such independencies conditions that are minimally sufficient for the same factor cannot be interpreted in terms of alternative causes of that factor. Accordingly, $A\bar{C}$ and $B\bar{C}$ cannot be alternative causes of W . This is straightforwardly mirrored by MT. The memberships of $A\bar{C}$ in MT_p and of $B\bar{C}$ in MT_q are not resistant against the introduction of further factors into the antecedents of MT_p and MT_q , respectively. Whenever $B\bar{C}$ is added to the antecedent of MT_p , $A\bar{C}$ is rendered redundant, and whenever the antecedent of MT_q is complemented by $A\bar{C}$, $B\bar{C}$ drops out for not being part of a minimally necessary condition. $A\bar{C}$ and $B\bar{C}$, mutually prevent each other from being causally interpreted by MT. Thus, even though both $A\bar{C}$ and $B\bar{C}$ are part of a minimal theory of W , none of them is attributed causal relevance by MT, for none of them remains part of the respective minimal theory upon factor frame extensions.

The chain-problem fundamentally differs from this minimalization problem. None of the factors in the antecedents of complex minimal theories as MT_n and MT_o violate (PR). Any of the disjuncts in these theories is instantiatable independently of all corresponding other disjuncts. If any actual factors happen to behave as described in (4.5), there cannot be any intuitions as to the underlying causal structure. In contrast, when confronted with specific interpretations of the factors in a set as R , causal intuitions are ready to hand. Whether the mechanism generating the movement of a car is chainlike or epiphenomenally structured is testable by extending factor frames. Therefore, contrary to the plural ways of minimalizing necessary conditions in complex double-conditionals, the ambiguities as regards the minimalization of necessary conditions in simple double-conditionals are unproblematic for our account of causation. The ambiguities simply exhibit the violation of (PR) due to conditional dependencies among minimally sufficient conditions of an investigated effect. There are extraordinary factor frames – as the one

³¹ Cf. section 2.4.

behind (4.5) – that a priori cannot conform to all causal principles. The behavior of the factors contained in such frames, which all feature conditional dependencies as in (4.6) and (4.7), cannot be causally modeled in principle.

4.8 Causal Reasoning and Coincidence Frames

In section 4.3.1 our discussion of the chain-problem started with the question as to whether, in principle, it is possible to identify causal structures by means of coincidence frames. Sentence (P), according to which different causal structures generate different coincidence frames, has been established as a sufficient condition for such an identifiability. The chain-problem, however, proves the invalidity of (P). In light of the solution to the chain-problem developed in this chapter, the question now arising, of course, is whether (P) is not only sufficient, but also necessary for the procedure of causal reasoning sketched in section 4.3.1. Does the invalidity of (P) prove the impossibility to infer causal structures from mere coincidence frames or is (P) modifiable in a way that nonetheless could serve as a ground on which to build a procedure of causal reasoning that operates on mere sets of coincidences?

(ICC) restores the unambiguous mapping of coincidence frames to causal structures within fully expanded factor frames. Entanglements that resist all factor frame extensions stem from causal dependencies among respective factors. That means, there are no entangled epiphenomena in fully expanded factors frames. Entanglements in epiphenomenal structures generally indicate the incompleteness of analyzed factor frames. In this sense, (ICC) can be seen as a restriction with respect to the well-formedness of causal structures. Nonetheless, relative to incomplete factor frames, it may well happen that different causal structures do not generate different coincidence frames. Graphs (n) and (o) on page 161 illustrate this. In light of (ICC), however, (P) can be modified to:

(P') Within fully expanded factor frames, different causal structures generate different coincidence frames.

According to (P') there may be identical coincidence frames resulting from different causal structures, yet only as long as corresponding factor frames are not fully expanded. Hence, (n) and (o) do not refute (P'). (P') merely states that upon expansions of the respective factor frame the set of coincidences compatible with a chain as (n) and with an epiphenomenon as (o) will diverge.

As a rule of thumb we introduced (CIE), which calls for a causal interpretation of entanglements, irrespective of whether factor frames are fully expanded or not. Upon expansions of particular factor frames this preference for the chain model may well turn out to be false. In fact, (P') guarantees that rash modelings will ultimately be corrected. Therefore, even though (P') is considerably weaker than (P), (P') still functions as a theoretical basis for a procedure that infers causal structures from coincidence frames. Such a procedure, however, will not be immune against (temporary) false causal diagnoses. Nonetheless, if (P') is valid, an unambiguous mapping of minimal theories onto causal structures is possible within

fully expanded factor frames. Against this background, coincidence frames would serve as empirical data based on which minimal theories and thereby, in a second step, causal structures could be derived. If (P') is valid, this chapter has thus developed the basic idea behind a procedure of causal reasoning that resorts to sets of coincidences as empirical data.

However, in light of the previous section, (P') might seem to be invalid, for section 4.7 has shown that an adoption of (ICC) does not exclude there being *mt*-equivalent minimal theories – even relative to fully expanded factor frames –, which, if causally interpreted, differ in causal respects. This is the case whenever minimally sufficient conditions of an effect are conditionally dependent in the sense of (4.6) and (4.7). Thus, whenever minimally sufficient conditions of an effect W imply a disjunction of other minimally sufficient conditions of W , necessary conditions of W are not unambiguously minimizable even within fully expanded factor frames. Does this amount to a violation of (P')? No, it does not, for a factor frame in which conditional dependencies along the lines of (4.6) and (4.7) hold violates the Principle of Relevance and, as a consequence thereof, dependencies among these factors are not causally interpretable. The dependencies among the factors contained in MT_p and MT_q are not differing *causal* dependencies generating identical coincidence frames, rather they are *logical* dependencies generating identical coincidence frames. MT_p and MT_q do not describe different causal structures that regulate the behavior of their factors identically.

Clearly, the fact that MT_p and MT_q do not rebut (P') does not prove the validity of (P'). (P') is a statement about the connection between causal structures and the sets of coincidences such structures generate. In this sense, it is a universally quantified empirical statement. As such, it defies rigid prove. It is acceptable as long as it has not been refuted by the discovery of different causal structures which – even in fully expanded factor frames – are compatible with the exact same coincidences. In what follows, the validity of (P') shall simply be assumed. The next chapter will flesh out the basic idea of a causal reasoning procedure as devised so far by developing an algorithm that assigns minimal theories to sets of coincidences. After having completed our conceptual analysis of causation, we thus now turn to causal reasoning.

5. ANALYSIS OF COMPLEX CAUSAL STRUCTURES

5.1 Introduction

This chapter develops a procedure of causal reasoning that infers causal structures of arbitrary complexity from the sets of coincidences these structures generate. Given the nature of its data input – sets of coincidences –, the procedure shall be termed *coincidence analysis*, or **CA** for short. **CA** is fully embedded in the theoretical framework of the regularity account of causation devised in the previous chapters. From the set of coincidences that are logically possible within a given factor frame minimal theories select the subset of actually instantiated, i.e. empirically possible, coincidences. This subset corresponds to the coincidence frame of a respective minimal theory.¹ Sets of empirically possible coincidences are thus assignable to minimal theories. The basic idea behind **CA**, therefore, consists in first assigning minimal theories to sets of coincidences and then causally interpreting these theories. This latter part of **CA**, of course, is straightforward. Minimal theories render causal structures syntactically transparent. Disjuncts in the antecedent represent alternative causes, conjuncts therein stand for parts of complex causes, and the factor in the consequent marks the effect. The difficult part of **CA** will be the first.

Minimal theories mirror the fundamental properties of causal structures, that satisfy the principles introduced in section 2.4, for instance: (a) causes are non-redundant parts of minimally sufficient conditions, which, in turn, are non-redundant parts of minimally necessary conditions; or (b) factors in complex causes are independent, i.e. they can be instantiated in all logically possible combinations; or (c) there are at least two alternative causes for each effect. These properties are reflected in sets of coincidences, e.g. as follows: If a set of coincidences does not contain two factors in all logically possible combinations, these factors cannot be part of the same complex cause, or if a set of coincidences contains no minimally sufficient conditions for a factor Z , Z must be a root factor.² Assigning minimal theories to coincidence sets essentially draws on identifying such structural properties in coincidence sets in a way that gradually narrows down the number of minimal theories that can be seen as properly representing the behavior of the factors in the corresponding set.

¹ Cf. section 4.2.

² Cf. p. 73 above.

Not any set of coincidences is causally analyzable and minimal theories are not always unambiguously assignable to coincidence sets.³ Our procedure of causal reasoning thus will not generally be applicable and it will at times infer multiple causal structures from a set of coincidences. (P'), however, guarantees that ambiguities will be resolved upon factor frame extensions. Such extensions do not always have to be fully completed, for not every causal investigation requires unambiguous causal diagnoses. Research interests often focus on causal dependencies among specific factors, while the structure regulating the remaining factors in a given frame is of secondary interest only.

CA builds on existing inference methodologies developed within the framework of modern regularity accounts.⁴ It differs from these existing methodologies in mainly two respects: (1) CA relies on considerably weaker causal assumptions and (2), in contrast to existing methodologies, it does not presuppose the subdivisibility of complex causal structures into their separate layers, but infers complex structures as a whole from coincidence data.⁵ Nonetheless, CA shares the core theoretical postulate with existing methodologies rooted in regularity accounts of causation: Causal relevancies can only be derived by comparison of *causally homogeneous* test situations. Therefore, before CA can be presented, the kernel of contemporary regularity theoretic procedures of causal reasoning must be reviewed.

5.2 Homogeneity

Causal reasoning consists of three types of inferences: diagnostic, prognostic, and theoretic inferences. The chapter at hand is concerned with the latter only. While in the course of diagnostic and prognostic inferences existing causal hypotheses are applied to investigated processes, it is the theoretic inference that generates new causal hypotheses.⁶ The theoretic inference is not only the form of causal reasoning that precedes diagnosis and prognosis, but also the by far most intricate type of causal inference. Diagnostic and prognostic inferences essentially apply universally quantified sentences – causal hypotheses – to particular cases. This application can be formalized by an unproblematic deductive inference. Yet, such as to account for the theoretic inference, as will be shown in this chapter, a considerably more elaborate technical apparatus is needed.

³ Analogous restrictions apply to modern probabilistic algorithms of causal reasoning as e.g. presented in Spirtes, Glymour, and Scheines (2000 (1993)).

⁴ Cf. e.g. Ragin (1987), May (1999), Ragin (2000) or Baumgartner and Graßhoff (2004).

⁵ Ragin (1987) has developed a methodology of causal analysis that meanwhile has become known as QCA (Qualitative Comparative Analysis) among social scientists. QCA bears considerable similarities to CA. It is a tool for the causal interpretation of pure *coincidence information* just as CA. Yet, in line with existing regularity theoretic inference procedures QCA presupposes the subdivisibility of complex structures into their layers. Furthermore, QCA is designed to analyze structures with a single effect only. In this sense, CA can be seen as a generalization of QCA. For details on the differences and commonalities of QCA and CA cf. Baumgartner (forthcoming).

⁶ Cf. Baumgartner and Graßhoff (2004), ch. VII.

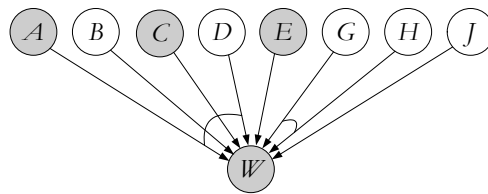


Fig. 5.1: Causal graph representing all factors causally relevant to W along with their structuring. The shaded vertices indicate the factors in a causally analyzed coincidence set.

Theoretic causal reasoning is brought to bear whenever the causal structure generating an investigated effect is unknown or partially known only. Effects result from complex causal nets, fractions of which – at most – are known to the person conducting causal analyses. A procedure of causal reasoning thus cannot straightaway identify the structure generating a particular effect as a whole. It starts its analysis by processing a manageable proper subset of all involved factors. If induced by the interests of a respective investigation, original factor frames are subsequently extended and further factors integrated into previously revealed parts of causal structures.

Suppose we are interested in the causes of an effect W . The graph in figure 5.1 shall be taken to represent the complete structure behind the behavior of W . Assume, the causal investigation starts without any previous knowledge about factors causally relevant to W . It shall only be established that events of type W regularly occur coincidentally with instances of the shaded factors A , C , and E . Accordingly, the shaded factors are the only ones incorporated into a causal investigation. The crucial task a procedure of causal reasoning now has to perform is to determine whether, say, A is causally relevant to W based on this incomplete subset of investigated factors. The mere observation of events a and w sequentially occurring is no evidence as to the causal relevance of A to W . Even a repeatedly correlated occurring of instances of A and W does not allow for the identification of causal relevancies. In every observed test situation, W could be brought about by instances of E , GH , or J without events of type A causally contributing at all. Moreover, A and W could be parallel effects of an as yet unknown common cause. Clearly though, the correlation might in fact be due to events of type A actually being parts of the causes of W . Answering the question as to the causal relevance of A to W hence is a problem of choosing between different possible causal hypotheses. Correlations of two factors may be the result of divergent causal structures. Figure 5.2 exhibits two causal structures that both generate a positive correlation of A and W .

Mill, who, as Hume, only accepted coincidences as causally analyzable data, proposed to solve this problem by imposing additional constraints on test situations:

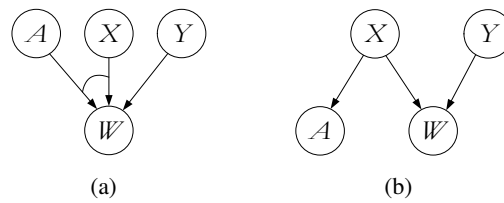


Fig. 5.2: Two divergent causal structures that both generate a correlation of A and W . As depicted in graph (a), A could be causally relevant to W . A and W might also be parallel effects of a common cause (cf. graph (b)).

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon.⁷

As is well known, this is what Mill referred to as the *method of difference*. According to Mill, if two test situations S_1 and S_2 can be brought about such that S_1 and S_2 only differ with respect to instantiations of A and W in S_1 , while in S_2 both of these factors are absent, the inference to the causal relevance of A to W is warranted. A is the only factor accountable for the instance of W in S_1 , for, apart from the presence and absence of W , an event of type A marks the only difference between S_1 and S_2 . Within such an experimental setting, a choice between the alternative hypotheses depicted in figure 5.2, in fact, is feasible. Given two test situations as S_1 and S_2 , only graph (a) can accurately represent the causal structure relating A and W . If the correlation of A and W were due to a structure as (b), there could not be two situations only differing in regard to A and W . Any test situation featuring an instance of W would moreover be constituted by events both of type A and X . Furthermore, whenever W were absent, both A and X would be absent as well.

The additional constraint that Mill imposes on causally analyzable test situations is commonly referred to as the *homogeneity condition*. According to Mill, two test situations S_1 and S_2 are causally homogeneous iff, apart from instances of the investigated effect, S_1 and S_2 differ in exactly one factor A , the so-called *test-factor*.

Homogeneity (I): In a causal test that investigates the causal relevance of a test-factor A and its negation to an effect W (or \bar{W}), two test situations S_1 and S_2 are homogeneous iff for all factors Z_i : Z_i is instantiated in S_1 iff Z_i is instantiated in S_2 , such that for Z_i :

$$Z_i \neq A, Z_i \neq \bar{A}, Z_i \neq W \text{ and } Z_i \neq \bar{W}.$$

⁷ Mill (1879 (1843)), p. 256.

This setting guarantees that, if the effect occurs upon an instantiation of A and is absent upon the absence of A , the inference to the causal relevance of A to W is warranted. If, however, W is only instantiated upon the absence of A , \bar{A} can be determined to be causally relevant to W .

Mill's version of homogeneity solves the problem of choosing between modeling a correlation of two factors in terms of a structure as depicted in graph (a) or in terms of a structure as exhibited by graph (b). Mere observations of sequentially occurring events – even if frequently repeated – do not back causal inferences unless they have been conducted against a very specific background such that evaluated test situations satisfy rigorous constraints. Yet, homogeneity (I) suffers from a severe flaw: It is virtually unsatisfiable. There simply are no pairs of different test situations in nature that, apart from instances of the investigated effect, coincide in all but one factor. Even in the course of an overly simple test, say, to clarify whether two chemicals interact, different test tubes are used or the test situations are sequentially realized, both of which violates homogeneity in Mill's sense. The test tubes differ as regards their atomic structure and their spatiotemporal properties. Homogeneity (I) is far too rigid. Though it could theoretically back causal inferences, a procedure of causal reasoning based on Mill's homogeneity would be inapplicable. Homogeneity thus needs to be weakened such that it can be satisfied by as many test situations as possible. While this weakening shall enhance satisfiability, it, of course, must not deprive homogeneity of its ability to back causal inferences and to decide between concurring causal hypotheses.

Such as to render test situations causally analyzable, they do not have to coincide in all but one factor and the effect. In order to manipulate a factor A , whose causal relevance is under investigation, numerous other factors among the causes of A may be required. As long as these 'manipulators' are not – as X in graph (b) of figure 5.2 – connected to the investigated effect W by a causal path on which A is not located, their causal relevance to W is always mediated by A . If such manipulators of A vary in test situations, the validity of inferences to the causal relevance of A to W cannot, under any circumstances, be compromised. For if graph (b) in figure 5.2 were to correspond to the actual causal structure underlying the behavior of A and W , there could not be two test situations differing exactly in W , A , and the latter's causes that are not located on any causal path entering W and not containing A . The factors causally relevant to both A and W , whose relevance to A is not mediated by W and whose relevance to W is not mediated by A , constitute the set of common causes of A and W .⁸ The factors that are causally relevant to a factor A , whose relevance to an effect W is being evaluated, and that do not belong to the common causes of A and W will be of considerable importance in the following. For brevity, we shall refer to them as *genuine causes* of the test-factor A or as *genuine test-factor causes*. That means, S_1 and S_2 are causally analyzable if, apart from A and W , they also differ with respect to genuine test-factor causes. Moreover, it is obvious that factors mediating the causal relevance

⁸ Cf. section 2.3.6 above.

of a test-factor to the investigated effect must be allowed to vary in test situations as well. If factors on a causal path from A to W , that are located between these two factors, are homogenized, a potential causal relevance of A to W cannot in principle show up, for then the behavior of W would be completely independent of whether A is present or not. Thus, factors that are located on a chain in between test-factors and effects – *intermediate factors* for short – are not to be homogenized in test situations.

Homogeneity (II): In a causal test that investigates the causal relevance of a test-factor A and its negation to an effect W (or \overline{W}), two test situations S_1 and S_2 are homogeneous iff for all factors Z_i : Z_i is instantiated in S_1 iff Z_i is instantiated in S_2 , such that for Z_i :

1. $Z_i \neq A$, $Z_i \neq \overline{A}$, $Z_i \neq W$ and $Z_i \neq \overline{W}$,
2. Z_i is no genuine test-factor cause and no intermediate factor between A and W .

Even though this constitutes a weakening of homogeneity (I), it still is hardly satisfiable. The aforementioned test as to the potential interaction of two chemicals is not only precluded from causal interpretability by (I), but also by (II). In addition to effects, test-factors, genuine test-factor causes, and intermediate factors any two test situations diverge with respect to their spatiotemporal coordinates. (II) has to be further attenuated:

In how much detail should we describe the situations in which this relation must obtain? We must include all and only the other causally relevant features.⁹

In this quote, Cartwright suggests a further weakening of homogeneity (II). Test situations do not have to coincide in all factors apart from test-factors, effects, genuine test-factor causes, and intermediate factors, but only in all factors *causally relevant* to the effect. This attenuation of (II) can draw on a certain amount of plausibility. If, for instance, the method of difference is applied in order to determine whether the striking of a match is causally relevant to its catching fire, it is of no importance whether the heads of the matches used in the two test situations are identically colored or whether one test situation is realized shortly before and the other directly after midnight. It suffices that both situations feature instances of the same factors causally relevant to matches catching fire, i.e. that there is oxygen present in both situations, that both matches are dry, and, especially, that in both situations alternative causes of matches catching fire are homogenized.

Homogeneity (III): In a causal test that investigates the causal relevance of a test-factor A and its negation to an effect W (or \overline{W}), two test situations S_1 and S_2 are homogeneous iff for all factors Z_i : Z_i is instantiated in S_1 iff Z_i is instantiated in S_2 , such that for Z_i :

⁹ Cartwright (1983), p. 29.

1. $Z_i \neq A$, $Z_i \neq \bar{A}$, $Z_i \neq W$ and $Z_i \neq \bar{W}$,
2. Z_i is no genuine test-factor cause and no intermediate factor between A and W ,
3. Z_i is *causally relevant* to W (or \bar{W}).

This third variant of homogeneity is noticeably weaker than the first two, and, accordingly, more easily satisfied. Nonetheless, homogeneity (III) guarantees for unambiguous causal inferences. Consider again the example depicted in figure 5.2. If two test situations can be realized that are homogeneous in the sense of (III), such that in one situation A and W are instantiated while in the other they are both absent, only graph (a) correctly represents the causal structure underlying the behavior of A and W .¹⁰

Still however, (III) imposes far-reaching constraints on causally analyzed test situations. Further attenuations are possible. In fact, the method of difference generates unambiguous results even if test situations do not coincide with respect to *all* causally relevant factors apart from test-factors, effects, genuine test-factor causes, and intermediate factors. As the previous chapters have shown, causally relevant factors are non-redundant parts of minimally sufficient conditions of their effects. In light of two test situations S_1 and S_2 such that S_1 features the investigated effect W , while S_2 does not, the causal relevance of a test-factor A can be established if it is merely guaranteed that the instance of W in S_1 is not caused by a minimally sufficient condition of W not containing A . The following fourth variant of causal homogeneity expresses this further weakening of (III).

Homogeneity (IV): In a causal test that investigates the causal relevance of a test-factor A and its negation to an effect W (or \bar{W}), two test situations S_1 and S_2 are homogeneous iff of all *minimally sufficient conditions* X_i of W (or \bar{W}) at least one conjunct is absent in S_1 iff at least one conjunct of X_i is absent in S_2 , where X_i satisfies the following conditions:

1. A , \bar{A} , W , \bar{W} are not part of X_i ,
2. no conjunct of X_i is a genuine test-factor cause or an intermediate factor between A and W ,
3. the conjuncts of X_i are *causally relevant* to W (or \bar{W}).

Provided that of all minimally sufficient conditions of W that comply with the requirements specified in (IV) at least one conjunct is absent in S_1 iff at least one conjunct is absent in S_2 , and given that W is only instantiated in S_1 – when A is present –, it follows that A is part of a minimally sufficient condition of W . For if W in S_1 were caused by a minimally sufficient condition X_i not containing A , such that X_i neither is a genuine test-factor cause nor an intermediate condition,

¹⁰ Accordingly, many authors favor this weakened version of causal homogeneity (cf. e.g. Johnson (1963 (1924)), vol. 3, or Blalock (1961), p. 22).

W would be instantiated in S_2 as well – given, of course, that S_1 and S_2 comply with homogeneity (IV). For in this case, X_i would be instantiated not only in S_1 , but also in S_1 's homogeneous counterpart S_2 . On the other hand, if an instance of W occurs in S_2 , while the homogeneous test situation S_1 does not feature such an instance, S_1 and S_2 impose an inference to the causal relevance of \bar{A} to W .

Therefore, two situations S_1 and S_2 that allow for an unambiguous inference to the causal relevance of a test-factor A (\bar{A}) to an effect W are not required to satisfy homogeneity (III). Homogeneity (IV) will do. (IV) does not impose constraints on instantiations of single factors. S_1 and S_2 may well feature different factors causally relevant to W – as long as every minimally sufficient condition X_i of W complying with the requirements imposed by (IV) is either instantiated in both test situations or in none. In the latter case, any conjuncts of X_i may be absent in S_1 and S_2 .

As long as only *two* test situations are compared, parts of minimally sufficient conditions containing the test-factor are allowed to vary arbitrarily.¹¹ This calls for a separate explication. Imposing compliance with homogeneity (IV) on two test situations serves the prevention of false causal conclusions, i.e. homogeneity (IV) is designed to back up unambiguous causal inferences. Thus, if there is an instance of the investigated effect in S_1 , yet not in S_2 , homogeneity must secure the validity of inferring the causal relevance of the test-factor. Such an inference cannot yield false conclusions upon variations of factors that are part of minimally sufficient conditions containing the test-factor. Suppose, AX constitutes a complex cause of W as depicted by graph (a) in figure 5.2. Moreover, assume two test situations S_1 and S_2 are observed that comply with homogeneity (IV) such that W is instantiated in S_1 , but not in S_2 . The absence of W in S_2 indicates that the instance of W in S_1 is caused by AX . That means, S_1 features both the test-factor A and X . In S_2 , however, A is suppressed, such that AX as whole is not instantiated, irrespective of whether X is present in S_2 or not. This constellation establishes the causal relevance of A to W , even if X varies in S_1 and S_2 . In contrast, if X is not instantiated in S_1 , while there is an instance of X in S_2 , W is absent in S_1 if and only if it is absent in S_2 as well – of course, assuming again that the causal structure behind the behavior of A and W corresponds to graph (a) in figure 5.2 and that homogeneity (IV) is satisfied. Either in both test situations alternative causes of W are instantiated, such that W is present in S_1 and S_2 , or in both test situations there are no instances of alternative causes, such that W is absent in S_1 and S_2 – in S_1 because X is missing and in S_2 because A is missing. In this latter case, nothing with respect to the causal relevance of A will be concluded. All this amounts to the following: If X is present in S_1 and absent in S_2 , homogeneity guarantees for an unambiguous inference to the causal relevance of A ; if, however, X is absent in

¹¹ We shall see shortly that a causal analysis of more than two test situations calls for a further specification of causal homogeneity to the effect that this variability of parts of minimally sufficient conditions containing the test-factor is prohibited.

S_1 and present in S_2 , no causal inference as regards a potential causal relevance of A is possible. Variations of X , hence, cannot induce false causal conclusions.

Furthermore, minimally sufficient conditions of W containing W itself are constituted by W as sole conjunct. W is itself minimally sufficient for W . Any conjunctive expansion of W by further factors would generate redundant conjuncts. Therefore, the first requirement imposed on X_i in (IV) to the effect that minimally sufficient conditions of W (or \overline{W}) containing W (or \overline{W}) itself are not to be homogenized amounts to the same as the requirement expressed in (I), (II), and (III) allowing test situations to vary with respect to instances of the investigated effect. Without such a specification, the method of difference would be inapplicable.

Of course, even homogeneity (IV) is far from easily satisfied. Such that test situations allow for unambiguous causal inferences, high standards have to be met. In light of our incomplete causal knowledge, there can never be certainty as to whether two test situations in fact are homogeneous in terms of (IV) or not. Nonetheless, only homogeneous test situations ensure unambiguous causal reasoning.

The question as to what premisses enter a theoretic causal inference and what standards causally analyzable test situations have to meet fundamentally differs from the question as to how test situations are evaluated with respect to their putative compliance with these standards. The first question is of theoretic or logical, the second of practical or epistemic nature. For the time being, we are concerned with theoretic requirements of causal inferences only. If merely two test situations are analyzed, homogeneity (IV) is an indispensable part of these requirements. Even if homogeneity of test situations may never be beyond doubt, assembling the theoretic preconditions of conclusive causal inferences, nonetheless, serves the goal of establishing experimental standards, implementable as a gauge for concrete experimental settings and causal inferences. In fact, there are ways to estimate the plausibility of two test situations conforming to homogeneity (IV). Such a homogeneity test has been developed in Baumgartner and Graßhoff (2004).¹²

For the purposes of the study at hand, homogeneity will be treated as an assumption that is presumed to be satisfied by test settings that are analyzed by our procedure of causal reasoning. This, of course, leaves room for false causal diagnoses. For whenever test situations are falsely assumed to be homogeneous, causal inferences are susceptible to errors. The latter, however, is a problem of epistemic nature only and does not affect the validity and conclusiveness of the procedure. Provided the premisses – one of which will be homogeneity of test situations – that enter the procedure are true, its output, i.e. its causal inferences, will necessarily be true as well. Yet, if one of the premisses is false, the output cannot be reliable. False inferences do not compromise the validity of the procedure, but merely indicate that one of the preconditions of a successful application of the procedure has not been satisfied. Or put differently: By making explicit the assumptions that have to enter causal inferences in order for them to be unambiguous, false conclusions

¹² Cf. Baumgartner and Graßhoff (2004), ch. 10.

are not prevented, rather, false conclusions are rendered precisely attributable to unsatisfied preconditions.

5.3 The State of the Art in Causal Reasoning within a Regularity Theoretic Framework

5.3.1 Method of Difference

Applying the method of difference as illustrated in the previous section yields a so-called *difference test*, or *df-test* for short. In the course of a df-test two situations are compared that comply to homogeneity (IV). A df-test evaluates the potential causal relevance of a test-factor A to an investigated effect W .

Ordinarily, causal tests are only brought to bear in view of processes that are assumed to be causally structured. Implementing procedures of causal reasoning without presuming the events under investigation to be the result of a causal structure would be pointless. That means, before a df-test is conducted it is assumed that some as yet unknown minimal theory of the investigated effect W exists. We express this unknown theory by means of variables:

$$(X_1 \vee Y_W) \Rightarrow W. \quad (H_1)$$

A df-test first of all generates an empirical datum: The effect W either does or does not occur upon the presence or absence of the test-factor. Such an empirical datum is commonly represented by so-called *coincidence tables* as depicted in table 5.1. The two fields correspond to two homogeneous test situations. Depending on whether the effect W is instantiated in a test situation or not, the corresponding field contains a '1' or a '0'. Accordingly, the two fields are referred to as *effect fields*. All in all, a df-test can generate $2^2 = 4$ different results. Table 5.2 lists these possible outcomes.

Not every df-test outcome is causally interpretable. Table (a) backs an inference to the causal relevance of A to W . In this table's first field the effect occurs, thus, at least one causally relevant minimally sufficient condition of W must be instantiated in the corresponding test situation. Necessarily, A is part of one of these conditions, for otherwise W would occur in the homogeneous test situation corresponding to field 2 as well. Table (b), on the other hand, induces an inference to the causal relevance of \bar{A} to W . Again, it is causal homogeneity that allows for this inference. Necessarily, \bar{A} is part of at least one causally relevant condition instantiated in field 2. For if that were not the case, there would be an event of type W in the test situation corresponding to field 1 as well. Tables (c) and (d) do not

D_0	A	\bar{A}
	field 1 (F1)	field 2 (F2)

Tab. 5.1: The results of df-tests are displayed by means of coincidence tables of this sort.

<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 2px 5px;">D_1</td><td style="padding: 2px 5px;">A</td><td style="padding: 2px 5px;">\bar{A}</td></tr> <tr><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">0</td></tr> </table> <p style="text-align: center; margin: 0;">(a)</p>	D_1	A	\bar{A}		1	0	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 2px 5px;">D_2</td><td style="padding: 2px 5px;">A</td><td style="padding: 2px 5px;">\bar{A}</td></tr> <tr><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px;">0</td><td style="padding: 2px 5px;">1</td></tr> </table> <p style="text-align: center; margin: 0;">(b)</p>	D_2	A	\bar{A}		0	1	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 2px 5px;">D_3</td><td style="padding: 2px 5px;">A</td><td style="padding: 2px 5px;">\bar{A}</td></tr> <tr><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">1</td></tr> </table> <p style="text-align: center; margin: 0;">(c)</p>	D_3	A	\bar{A}		1	1	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 2px 5px;">D_4</td><td style="padding: 2px 5px;">A</td><td style="padding: 2px 5px;">\bar{A}</td></tr> <tr><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px;">0</td><td style="padding: 2px 5px;">0</td></tr> </table> <p style="text-align: center; margin: 0;">(d)</p>	D_4	A	\bar{A}		0	0
D_1	A	\bar{A}																									
	1	0																									
D_2	A	\bar{A}																									
	0	1																									
D_3	A	\bar{A}																									
	1	1																									
D_4	A	\bar{A}																									
	0	0																									

Tab. 5.2: Possible outcomes of a df-test investigating the causal relevance of a test-factor A to an effect W . Table (a) induces an inference to the causal relevance of A to W . From table (b) the causal relevance of \bar{A} to W can be inferred. Tables (c) and (d) do not give rise to any causal inferences.

back any causal inferences. While the test situations represented by table (c) apparently feature minimally sufficient conditions neither containing A nor \bar{A} , such that nothing with respect to a possible causal relevance of A or \bar{A} can be determined, the impossibility to draw causal inferences based on table (d) may not be as obvious. Prima facie it might be held that (d) authorizes an inference to the *causal irrelevance* of A or \bar{A} to W . That, however, is not the case. Homogeneity (IV) does not require all factors possibly constituting a complex cause of W in combination with A or \bar{A} to be instantiated in test situations. Hence, if a df-test yields the result depicted in table (d), the causal irrelevance of A or \bar{A} is by no means established. Assume, A is part of a complex cause X of W . Notwithstanding this fact, W could be absent in the situation represented by field 1, simply because some other factor of X is absent in that situation.

Clearly, if homogeneous test situations were required to lack *exactly one* factor of each complex cause of W , the df-test outcome reported in table (d) would induce an inference to the causal irrelevance of both A and \bar{A} to W . For provided that the effect is absent in two test situations that are homogeneous in this more restrictive sense, neither A nor \bar{A} could be held to complete a complex cause of W . Therefore, neither A nor \bar{A} would be part of a complex cause of W . They would thus both be causally irrelevant to W . The problem with such a strengthening of homogeneity, however, is at hand: The more restrictive homogeneity, the harder its experimental realization. In light of our sweeping causal ignorance, it is hardly realistic to ask of test situations to lack exactly one factor of each complex cause.¹³ With regard to keeping the presumptions entering our procedure of causal reasoning as weak as possible and thereby facilitating its application, we shall abstain from such a strengthening. The cost of this abstinence is that our procedure will not be able to infer causal irrelevancies. Apart from illustrating the pros and cons of homogeneity designed along the lines of (IV), the above considerations, hence, demonstrate that detecting causal relevancies imposes far weaker constraints on test situations than the identification of irrelevancies.

Given that a df-test generates the tables (a) or (b) of figure 5.2, the original causal hypothesis (H_1) is expandable. In light of table (a), the minimal theory (H_1) can be adapted to:

$$AX'_1 \vee Y_W \Rightarrow W. \tag{H_2}$$

¹³ In section 5.5.2, on page 226, we shall encounter a further reason as to why a procedure of causal reasoning that operates on coincidence data only cannot infer causal irrelevancies.

Table (b), in turn, allows for complementing (H₁) as follows:

$$\overline{A}X_1' \vee Y_W \Rightarrow W. \quad (\text{H}_3)$$

The variable X_1 in (H₁) is tagged with a stroke in (H₂) and (H₃), for its domain is being modified upon the expansion of (H₁). In both (H₂) and (H₃) one conjunct contained in the domain of X_1 in (H₁) is picked out and rendered explicit. This expansion of (H₁) by A and \overline{A} modifies the domain of X_1 .

5.3.2 The four-field test

Given that the outcome of a simple df-test corresponds to D₁ or D₂ (tables (a) or (b) in figure 5.2), the method of difference allows for inferences to causal relevancies of single factors. In this way, df-tests identify classes of causally relevant factors. A df-test, however, cannot determine whether pairs of causally relevant factors are part of the same complex cause or whether the elements of such pairs are contained in alternative causes. That means, the method of difference cannot structure classes of causally relevant factors. The reason for this limited analytical power is the following: Based on a df-test, two factors A and B are only attributable causal relevance to an investigated effect W *independently* of each other. A df-test can establish that A and B are each part of a causally relevant minimally sufficient condition of W , yet whether these conditions coincide or not is a question a df-test cannot answer. In order to structure classes of causally relevant factors, i.e. in order to infer complex and alternative causes, May (1999) introduced an extended experimental setting: the so-called *four-field* test or *ff-test* for short.¹⁴

Except for homogeneity as formulated in (IV), the same assumptions enter the ff- as the df-test. Clearly though, test situations constituting an ff-test need to be causally homogeneous as well. (IV), however, only defines homogeneity for two test situations – one with an instance of the test-factor, one without. In contrast, an ff-test not only evaluates the causal relevance of a test-factor B , but, if B is found to be causally relevant, an ff-test also determines whether B is part of the same complex cause as an already identified causally relevant factor A or whether A and B belong to different alternative causes. For these purposes, a systematic variation of A and B is required, which, in case of two test-factors, calls for four test situations. While in test situations of a df-test factors that are part of the same complex cause as the test-factor are allowed to vary arbitrarily¹⁵, such *co-factors* of the test-factors must be homogenized in an ff-test. Suppose, $ABCD$ is a complete complex cause of an effect W . The test situations analyzed in the course of an ff-test investigating the test-factors A and B are causally homogeneous only if *the rest* of $ABCD$, i.e. CD , is either instantiated or absent in *every* test situation, in which one of the test-factors is instantiated. That means, in the course of our

¹⁴ That a structuring of causally relevant factors is impossible when cause-effect pairs only are analyzed has been forcefully pointed out by Ragin (1987). His QCA-methodology, accordingly, calls for a *holistic* perspective when it comes to uncovering causal structures.

¹⁵ Cf. p. 198 above.

exemplary ff-test, instances of A and B occur only in combination with either instances of CD or with instances of $C\bar{D}$, $\bar{C}D$ or $\bar{C}\bar{D}$. Without such an additional homogenizing of co-factors, the structuring of the causal process at hand might turn out to be fallacious. If co-factors are not homogenized, it is not possible to unambiguously determine whether A and B – if attributed causal relevance – are to be seen as parts of the same minimally sufficient condition of W or not.¹⁶ (HC*) amends homogeneity (IV) in this sense.

Homogeneity (HC)*:¹⁷ In a causal test that investigates the causal relevance of the test-factors A , \bar{A} , B , and \bar{B} to an effect W (or \bar{W}), four test situations are homogeneous iff

- (1) of all *minimally sufficient conditions* X_i of W (or \bar{W}) at least one conjunct is absent in one of the test situations iff at least one conjunct of X_i is absent in all other three test situations as well, where X_i satisfies the following conditions:
 1. A , \bar{A} , B , \bar{B} , W , and \bar{W} are not part of X_i ,
 2. no part of X_i is a genuine test-factor cause or an intermediate factor between a test-factor and W ,
 3. the parts of X_i are *causally relevant* to W (or \bar{W}).
- (2) the rest of a complex cause X_j of W (or \bar{W}) such that A , \bar{A} , B , or \bar{B} are part of X_j , does *not vary* in test situations featuring the test-factor(s) contained in X_j .

The results of ff-tests are commonly displayed by tables as 5.3. The fields 1 to 4 represent test situations that are homogeneous in terms of (HC*). As in df-test tables, ‘1’ indicates an instantiation of the investigated effect and ‘0’ its absence.

Commonly, an ff-test is preceded by a df-test that has generated a first causally interpretable result. Such as to illustrate this sequential ordering of ff- and df-tests, let us resume the example discussed in the previous section. Suppose, a successful df-test has determined the causal relevance of a factor A to an effect W and has thus induced expanding (H_1) in terms of (H_2). Prior to conducting an ff-test, hence,

V_0	B	\bar{B}
A	field 1 (F1)	field 2 (F2)
\bar{A}	field 3 (F3)	field 4 (F4)

Tab. 5.3: The results generated by ff-tests are represented by tables of this form. A and B are the systematically varied test-factors.

¹⁶ Cf. page 212 below.

¹⁷ (HC*) is tagged with a “*” because it does not explicate causal homogeneity in the generality that will be needed for our procedure (CA) below. (HC*) thus is provisional and will be further amended in section 5.5.3.

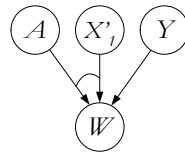


Fig. 5.3: Causal graph representing the causal structure behind the behavior of an effect W . A is the only known causally relevant factor.

(H₂) is taken to be established. Given the causal relevance of A , we now want to know whether a second factor B is causally relevant to W as well and, if that is the case, where in (H₂) B has to be integrated.

The graph in figure 5.3 depicts the established causal knowledge concerning W relative to the just delineated epistemic situation. Factor A is causally relevant to W . Yet, most likely, A is not itself minimally sufficient for W . A is part of at least one complex condition AX'_1 , which, as a whole, is not necessary for W . The incomplete graph in figure 5.3, thus, is expandable in manifold ways. For example, further factors could be integrated into condition AX'_1 or there is room for alternative causes X_2 or for intermediate factors that are located on a causal chain from AX'_1 to W or, finally, for factors that are causally relevant to A and, accordingly, indirectly relevant to W .

Therefore, the question that a continued analysis of the causal structure behind the behavior of W has to answer is whether the second test-factor B has to be added

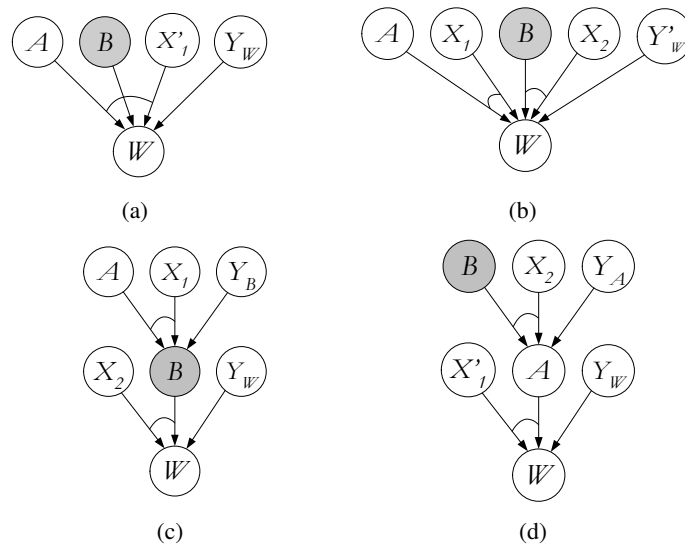


Fig. 5.4: Four possibilities to integrate factor B , provided it is causally relevant to W , into the graph of figure 5.3.

to the graph of figure 5.3 and, if that is the case, where its proper place is. Given that B is found to be causally relevant to W , there are four possible expansions of our exemplary causal graph.

- (a) B could be conjunctively added to the minimally sufficient condition already containing A .
- (b) B could be part of an alternative minimally sufficient condition not yet explicitly represented in graph 5.3. In this case, B would be the first known factor of that alternative cause.
- (c) B could be located between AX'_1 and W on a chain such that A 's causal relevance to W would turn out to be of indirect nature.
- (d) B could be causally relevant to A and, thus, merely indirectly relevant to W by mediation of A .

These four possible expansions of the causal structure depicted in figure 5.3 are graphed in figure 5.4.

Causal Independence of Test-Factors

This subsection shall show that distinguishing between the four graphs in figure 5.4 by means of a four-field test calls for the adoption of a further assumption, because the data generated by an ff-test empirically underdetermines causal inferences. More specifically: There are certain outcomes of ff-tests that do not unambiguously induce one of the four possible integrations of a second test-factor into an already established causal structure. In order to illustrate this problem, consider the ff-test result depicted in table (a) of 5.4. Within a (HC*)-homogeneous background, the effect W is instantiated if and only if B is present. The difference between the fields F1 and F2 or between F3 and F4 establishes the causal relevance of B to W . Yet, it is undetermined where B is to be integrated into an existing structure constituted by A as only known causal factor. Table (b) in 5.4

			fields of V_e	homogeneous coincidences
			F1	$\overline{AX_1Y_B}BX_2\overline{Y_W}W$
			F2	$\overline{AX_1Y_B\overline{B}X_2\overline{Y_W}W}$
			F3	$\overline{AX_1Y_B}BX_2\overline{Y_W}W$
			F4	$\overline{AX_1Y_B\overline{B}X_2\overline{Y_W}W}$
V_e	B	\overline{B}		
A	1	0		
\overline{A}	1	0		
(a)			(b)	

Tab. 5.4: The ff-test table V_e in (a) establishes the causal relevance of B to W , yet does not allow for determining whether A and B are contained in different complex causes of W or whether they constitute a chain. Table (b) lists four background scenarios possibly constituting the test situations of V_e . The rows of (b) correspond to the fields of V_e . Each row indicates the causal background of the corresponding field.

lists four background scenarios that could possibly generate the ff-test outcome V_e such that each row corresponds to one of the fields in V_e . These background scenarios could both stem from a one-layer and from a two-layer structure. The coincidence set listed in table (b) of 5.4 is compatible with both of the following minimal theories, which correspond to graphs (b) and (c) in figure 5.4:

$$AX_1 \vee BX_2 \vee Y_W \Rightarrow W \quad (5.1)$$

$$(AX_1 \vee Y_B \Rightarrow B) \wedge (BX_2 \vee Y_W \Rightarrow W) \quad (5.2)$$

In order to demonstrate this underdetermination, first assume (5.1) to be the true causal structure regulating the behavior of W . Since X_1 is permanently absent in the four test situations of V_e , there is no instance of W in F2. This finding, however, does not threaten the causal relevance of A to W , for, as mentioned above, this relevance is presumed to have been established in the course of a foregoing df-test. In F1 and F3 the events of type W are caused by instances of BX_2 . Taken together, this shows that the scenarios in table (b) can be seen as stemming from a structure such that A and B are contained in alternative causes of W as expressed by (5.1). Second, suppose (5.2) to be the causal structure behind V_e . Due to the permanent absence of X_1 , A cannot demonstrate its already established (indirect) causal relevance in V_e . Accordingly, B is not manipulated by means of AX_1 , but by some as yet unspecified alternative cause in Y_B . Whenever B is instantiated, the effect occurs. This indicates that the coincidence set displayed in table (b) might as well be the result of a causal chain such that A is a direct cause of B and B is a direct cause of W – as stated by (5.2).

This finding, of course, does not amount to the *mt*-equivalence of (5.1) and (5.2). The coincidence set in table (b) of 5.4 is only a small subset of all 128 coincidences that are logically possible within the factor frame at hand. For instance, suppose that X_1 – the co-factors of A – were permanently instantiated instead of absent as in table (b) and that W were instantiated in F2 as well. This constellation would unambiguously induce an inference to the one-layer model. For there would no longer exist a homogeneous scenario compatible with the chain represented in (5.2). $AX_1\overline{Y_B}\overline{B}X_2\overline{Y_W}W$ is compatible with (5.1), yet not with (5.2). The two minimal theories thus are not *mt*-equivalent.

Nonetheless, however, the existence of at least one homogeneous background scenario of V_e possibly stemming from two different causal structures prohibits an inference to either the one- or the two-layer model. In order to allow for unambiguous causal inferences drawn from ff-tests, an additional causal assumption is introduced in May (1999) and Baumgartner and Graßhoff (2004): It is assumed that test-factors analyzed in four-field tests are *causally independent*.¹⁸

Causal independence of test-factors (IND): The test-factors systematically varied in the course of ff-tests are causally independent of each other.

¹⁸ Ragin's (1987) procedure of causal reasoning likewise assumes independence of test-factors.

Applied to our exemplary case, (IND) renders B introducible into the existing causal structure as root factor only. By assumption, B can neither be causally relevant to A nor causally dependent on A . B can only be introduced into the same layer as A . This excludes (5.2) as a structure possibly underlying V_e .

(IND) is necessary for drawing conclusive causal inferences from four-field tables. However, due to (IND) traditional inference procedures based on regularity theories are forced to analyze causal structures layer by layer. Single layers only are accessible to ff-tests. Assuming the set of test-factors analyzed in the course of an ff-test to consist of causally independent factors only, of course, is a far-reaching causal supposition. A number of causal structures possibly underlying a respective ff-test result are excluded merely by fiat. No doubt, it is a fact of experimental practice that test-factors are often known to be independent prior to conducting ff-tests. Nonetheless, (IND) is a causal assumption to the effect that important aspects of causal processes submitted to causal analysis by means of ff-tests have to be known prior to conducting ff-tests, which, in turn, should be designed to reveal just those aspects of the processes under investigation. One of the crucial features of *coincidence analysis* (CA) to be developed in this study will therefore consist in abandoning (IND).

Inference Rules

More will be said about (IND) and its consequences below. For now, we just take (IND) for granted and proceed by introducing the inference rules as proposed in May (1999) and Baumgartner and Graßhoff (2004). (IND) entails that the test-factors involved in an ff-test can be co-instantiated in all logically possible combinations. This yields $2^4 = 16$ possible ff-test outcomes. In Baumgartner and Graßhoff (2004) these 16 cases are discussed systematically along with the causal inferences they induce.¹⁹ In the present context, we can thus dispense with a full recapitulation of all 16 ff-test outcomes. I shall just pick out a few exemplary cases that suffice to introduce the pertinent inference rules.

Again, we resume the example initiated in section 5.3.1. Assume that a df-test has established the causal relevance of a first test-factor A to an effect W . We now look at five possible ff-test outcomes that each induce different expansions of this antecedently established causal structure (H_2).

$$1. \begin{array}{c|c|c} V_1 & B & \bar{B} \\ \hline A & 1 & 0 \\ \hline \bar{A} & 0 & 0 \end{array}$$

$$\text{Inference: } ABX_1'' \vee Y_W \Rightarrow W$$

The first column of V_1 accounts for the causal relevance of A , the first row establishes the same for the second test-factor B . If one of the two test-factors is

¹⁹ Cf. Baumgartner and Graßhoff (2004), pp. 217-222.

missing, no effect occurs. Since within this homogeneous setting the only relevant difference among the four test situations consists in the presence and absence of A and B , V_1 moreover demonstrates that A and B must be part of the same minimally sufficient condition accountable for the instance of W in F1. Yet, V_1 does not shed light on whether AB is a complete minimally sufficient condition of W or not. In every test situation of V_1 further unknown factors could be present that are contained in the same minimally sufficient condition as A and B . The condition AB in particular and the minimal theory inferred from V_1 in general, thus, have to remain open for subsequent expansions. As usual, this openness is expressed by means of the variable X_1'' running over additional conjuncts of the minimally sufficient condition containing A and B .

V_2	B	\bar{B}
2. A	0	0
\bar{A}	1	0

$$\text{Inference: } AX_1' \vee \bar{A}BX_2 \vee Y_W' \Rightarrow W$$

Again, factors A and B constitute the only possibly relevant difference among the four fields. The effect only occurs in the test situation corresponding to field 3. Thus, relative to the homogeneous background of V_2 , W is instantiated if and only if \bar{A} and B are co-instantiated. Therefore, \bar{A} and B must be non-redundant parts of one and the same minimally sufficient condition accountable for W in F3. The already diagnosed causal factor A and its negation, for logical reasons, cannot be part of this newly established minimally sufficient condition. Moreover, in F3 a minimally sufficient condition of W must be present that does not contain A , for the only previously known cause AX_1' is absent in F3. That amounts to introducing a second minimally sufficient condition of W besides AX_1' . The latter, in turn, is not deprived of its causal relevance to W in light of V_2 , even though W is not instantiated in F1. The absence of W in F1 merely indicates that at least one of the factors contained in AX_1' must be missing in V_2 . As in the case of V_1 , V_2 does not determine whether the newly introduced condition containing \bar{A} and B is fully specified yet. In order for this alternative cause to remain open for further expansions, the variable X_2 is introduced. The first disjunct of the minimal theory inferred from V_2 , however, does not follow from V_2 , but from the previously established minimal theory (H_2). (H_2) enters the causal inference drawn from V_2 as an implicit premiss, so to speak. Its causal relevancies can be transferred to the conclusion drawn from V_2 .

V_3	B	\bar{B}
3. A	1	1
\bar{A}	1	0

$$\text{Inference: } AX_1' \vee BX_2 \vee Y_W' \Rightarrow W$$

In V_3 at least two alternative causes of W are operating. The first row indicates that there is a causally relevant minimally sufficient condition of W that does not contain B , for W occurs irrespective of whether B is instantiated or not. For analogous reasons, the first column reveals that there is an alternative cause of W not containing A . B must be part of this second minimally sufficient condition, for the difference between F3 and F4 establishes the causal relevance of B . All in all, A and B must be part of different complex causes of W , whose incompleteness again is symbolized by variables.

Not all 16 possible outcomes of an ff-test are causally interpretable. The following two ff-test tables do not induce any expansion of the original causal hypothesis (H_2).

$$4. \begin{array}{c|cc} V_4 & B & \bar{B} \\ \hline A & 0 & 0 \\ \hline \bar{A} & 0 & 0 \end{array}$$

$$5. \begin{array}{c|cc} V_5 & B & \bar{B} \\ \hline A & 1 & 1 \\ \hline \bar{A} & 1 & 1 \end{array}$$

The reason for the impossibility to causally interpret these two ff-test tables is obvious: Whenever there are no differences between test situations, no differences are to be accounted for by attributing them the varying test-factors. In V_4 the effect does not occur because no minimally sufficient conditions of W are instantiated in any of the test situations. In V_5 the effect is permanently present because the homogeneous background features a minimally sufficient condition of W that does neither contain A nor B . Both cases do not shed any light on potential causal relevancies of B or \bar{B} . That means, these two ff-test tables neither demonstrate the causal relevance nor the causal irrelevance of the test-factors to the investigated effect.

An ff-test only investigates the interplay of two test-factors with respect to one effect. This, of course, still is a very restricted test setting, recourse to which is only possible at the very beginning of a causal investigation. After having expanded a minimal theory by two factors, it by far is not fully specified yet. As soon as more than two test-factors are analyzed, ff-tests will no longer do. However, the basic idea behind the ff-test set-up – systematic variation of all test-factors – is straightforwardly expandable to more than two test-factors. For an *extended causal test* involving n test-factors 2^n test situations are required, which are readily represented by tables consisting of 2^n fields. Thus, besides computational complexity, nothing methodologically new is introduced in extended causal tests. A systematic variation of (the causally independent) test-factors against a homogeneous background is both sufficient and necessary to attribute causal relevance to test-factors

and to determine their function and location within a causal structure. Reviewing this basic idea suffices for the purposes of the present study. The extended causal test setting will thus not be further elaborated here. Details can be found in May (1999) or Baumgartner and Graßhoff (2004).²⁰ For brevity's sake, I shall refer to this methodology of causal analysis as the *2ⁿ-method* in the following.

The causal inferences induced by ff-test tables and by tables generated by extended causal tests are regulated by the exact same inference rules. Given that causally independent test-factors are systematically varied with respect to presence and absence against a homogeneous background, one-layer causal structures are identifiable by means of three simple rules. The exemplary ff-test tables discussed above suffice to introduce these rules.

Rule of Difference A simple df-test compares two homogeneous test situations S_1 and S_2 . If the investigated effect W occurs upon the presence of the test-factor A (S_1) and is absent upon \bar{A} (S_2), it follows first of all that at least one minimally sufficient condition of W is instantiated in S_1 , while in S_2 that is not the case. Moreover, since upon satisfaction of homogeneity S_1 only differs from S_2 with respect to minimally sufficient conditions of W that contain the test-factor or with respect to genuine test-factor causes and intermediate factors, the absence of W in S_2 determines that in S_1 *no* minimally sufficient condition of W is instantiated that *does not* contain A or that *does not* constitute a genuine test-factor cause or that *is not* composed of intermediate factors located between A and W . At least one of the minimally sufficient conditions of W present in S_1 must be causally interpretable, for – due to the Principle of Causality – W is not instantiated without any of its causes. Genuine test-factor causes and intermediate factors are minimally sufficient for W only if S_1 additionally features a causally interpretable minimally sufficient condition of W containing A . All minimally sufficient conditions of W that are present in S_1 and that neither are genuine test-factor causes nor composed of intermediate factors contain A . It follows either (a) that at least one of the minimally sufficient conditions of W present in S_1 is a genuine test-factor cause or constituted of intermediate factors or (b) that all minimally sufficient conditions of W present in S_1 contain A . (a) only is the case if S_1 additionally features at least one causally relevant minimally sufficient condition of W containing A . In contrast, if (b) is the case, at least one of the minimally sufficient conditions of W present in S_1 is causally interpretable and contains A . Hence, irrespective of whether (a) or (b) holds, A is part of at least one causally interpretable minimally sufficient condition of W and, thus, is itself causally relevant to W .

We have derived the first and most fundamental rule of causal reasoning within a regularity theoretic framework: the so-called *rule of difference*.

Rule of Difference (RD): For a test-factor A , two homogeneous test situations S_1 and S_2 , and an effect W : If there is both an instance of A and of W in S_1 ,

²⁰ Cf. May (1999), pp. 216-224, or Baumgartner and Graßhoff (2004), pp. 228-231; also: Ragin (1987) and Ragin (2000).

while S_2 features neither an event of type A nor an event of type W , at least one causally relevant minimally sufficient condition of W containing A is instantiated in S_1 .

Based on a df-test \mathcal{RD} can establish the causal relevance of a single factor A . The combined instantiation of an effect W , of at least one of its minimally sufficient conditions, and of A is not sufficient evidence as to the causal relevance of A to W . The latter can only be inferred if it is additionally guaranteed that S_1 does not feature any minimally sufficient condition of W that does not contain A or that is not a genuine test-factor cause or that is not composed of intermediate factors located between A and W . In order to guarantee for this, a second test situation S_2 is required that is homogeneous to S_1 and that features neither an event of type A nor of type W .

Rule of Combination A procedure of causal reasoning not only has to identify factors that are causally relevant to an investigated effect. Over and above the simple identification of relevancies, it has to be able to determine whether pairs of causally relevant factors are part of the same complex cause or not. We are not only interested in atomic relevancies, but in complex causal structures. The inference rule that – by presumption of (IND) – allows for assigning causally relevant factors to the same complex cause is the so-called *rule of combination*. Such as to systematically develop that rule, let us reconsider the ff-test table V_1 .

V_1	B	\bar{B}
A	1	0
\bar{A}	0	0

The fields F1 to F4 represent four (HC*)-homogeneous test situations. The effect only occurs in F1. Thus, only F1 features a causally relevant minimally sufficient condition of W . In light of the absence of the effect in F2, \mathcal{RD} allows for an inference to the instantiation of at least one causally relevant minimally sufficient condition of W in (the situation corresponding to) F1 that contains B . Moreover, the absence of the effect in F3 induces an inference to the instantiation of at least one causally relevant minimally sufficient condition of W in F1 that contains A . Based on (IND) it is furthermore guaranteed that neither A is causally relevant to B nor vice versa. The only remaining question, hence, is whether (a) A and B are part of the same complex cause of W or (b) whether they belong to two alternative causes. That (a) is the case can indirectly be shown by demonstrating that (b) cannot hold in light of V_1 . Assume that F1 features two alternative causes AX_1 and BX_2 of W . If that were the case, the presumed satisfaction of (HC*) by V_1 would determine that W is instantiated upon absences of A and B , i.e. in the situations corresponding to F2 and F3. That, however, does not hold for V_1 . Therefore, A and B must be part of the same complex cause.

At this point it becomes apparent why causal homogeneity in terms of (IV) would not do for ff-tests or extended causal tests, i.e. why condition (2) had to be added in (HC*). If it were not guaranteed that the remaining factors of a minimally sufficient condition of W containing the test-factor are homogenized in F1 and F3, case (b) would also be compatible with V_1 . In this case, B 's location within the causal structure regulating the behavior of W would not be determinable.

Rule of Combination (RC): For any set of factors $\{A_1, \dots, A_n\}$, $1 \leq n$, any set of homogeneous test situations $\{S_1, \dots, S_{2^n}\}$, and any effect W : If in one of the test situations S_i the coincidence $A_1 A_2 \dots A_n W$ is instantiated and it moreover holds that

- A_1 to A_n are mutually causally independent according to (IND) and
- no other test situation S_j , $j \neq i$, features the coincidence $A_1 \dots \overline{A_k} \dots A_n W$, $1 \leq k \leq n$,

then A_1 to A_n are part of the *same* causally relevant minimally sufficient condition of W .

Hence, if in one of four homogeneous test situations of an ff-test there is both an instance of a causally relevant minimally sufficient condition of W containing A and an instance of such a condition containing B , and none of the test situations feature $A\overline{B}$ or $\overline{A}B$ in combination with W , it can be inferred – presupposing (IND), of course – that A and B are part of the same causally relevant minimally sufficient condition. \mathcal{RD} guarantees that F1 features a causally relevant sufficient condition of W containing A and B . The question answered by comparing F1 to F2 and to F3 is whether that sufficient condition is *minimal* or not. If the sufficient condition reduced by one conjunct still is sufficient for W , W is instantiated in F2 and F3 as well. If that is not the case, A and B must be part of the same *minimally* sufficient condition.

Rule of Alternatives \mathcal{RC} constitutes a means to identify complex causes based on the data generated by an ff-test or an extended causal test. What is still missing now is a rule that determines under what conditions test-factors whose causal relevance is established by \mathcal{RD} are to be assigned to alternative causes. This is just what is accomplished by the so-called *rule of alternatives*.

Rule of Alternatives (RA): If an effect W is instantiated in a test situation S_1 which does not feature any of its previously known causally relevant minimally sufficient conditions, there must exist at least one further causally relevant minimally sufficient condition of W .

According to the Principle of Causality, every effect has a cause. Therefore, if an effect is instantiated, notwithstanding the fact that all of its known causes are absent, the corresponding instance of the effect must have been caused by some

further, as yet unknown alternative cause. Such as to illustrate the application of this third inference rule, let us reconsider V_2 . Assume a df-test has established the causal hypothesis (H_2). Now, an ff-test generates the following result:

V_3	B	\bar{B}
A	0	0
\bar{A}	1	0

In F3 the effect W is instantiated without any of its previously known causes. For, by assumption, antecedently to conducting the ff-test under consideration only the causal relevance of A had been established. A , however, is missing in F3 and thereby every previously known cause of W . According to \mathcal{RA} , the instantiation of W in F3 must be due to a previously unknown alternative cause. By \mathcal{RC} it moreover follows that \bar{A} and B must be part of that previously unknown cause. A sequential implementation of \mathcal{RA} and \mathcal{RC} thus induces an expansion of the original hypothesis (H_2) to $AX'_1 \vee \bar{A}BX_2 \vee Y'_W \Rightarrow W$.

5.3.3 Modularity

These three inference rules constitute the core of the causal reasoning procedure as proposed in May (1999) and Baumgartner and Graßhoff (2004). By means of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} , each of the 16 possible outcomes of an ff-test and any result generated by extended causal tests are causally analyzable. The application of the rules either yields an expansion of an antecedently established causal hypothesis (minimal theory) or it reveals that such an expansion is not warranted by a respective test result – as e.g. tables V_4 and V_5 illustrate. It is important to note that the application of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} is modularly structured in twofold ways. First, due to (IND), the three rules can only be applied to single layers of causal structures. As soon as (IND) is violated, i.e. as soon as there are causal dependencies among test-factors, the rules no longer yield unambiguous causal diagnoses. Thus, their application presupposes a subdivisibility of multi-layer structures into single layers. Second, single layers of causal structures are analyzed by applying the three rules in an appropriate order. That means, only a suitably combined implementation of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} accomplishes a complete causal analysis of test tables.

While the limitation of the rules to one-layer structures will be more extensively discussed in the next section, the sequential implementation of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} shall be further illustrated here. Assume that, after having confirmed the validity of (H_2) by means of a df-test, an ff-test conducted to detect further causes of the effect W generates to following data:

V_6	B	\bar{B}
A	0	0
\bar{A}	0	1

By comparison of the fields F3 and F4, \mathcal{RD} establishes the causal relevance of \overline{B} to W . Likewise based on \mathcal{RD} , causal relevance can be attributed to \overline{A} upon a comparison of F2 and F4. Only F4 features an occurrence of the effect. By implementation of \mathcal{RC} , it can thus be derived that \overline{A} and \overline{B} are part of the same causally relevant minimally sufficient condition $\overline{AB}X_2$ of W . Moreover, since the only previously known causal factor A is missing in F4, the instantiation of W in F4 must be ascribed to another, as yet unknown alternative cause. Accordingly, \mathcal{RA} allows for inferring that A and \overline{AB} are contained in different alternative causes of W . All this amounts to the following expansion of the initial hypothesis (H_2):

$$AX'_1 \vee \overline{AB}X_2 \vee Y'_W \Rightarrow W. \quad (5.3)$$

Tabulating these separate inference steps yields the following proof scheme for (5.3):

INFERENCE	SANCTION	RULE
(1) \overline{B} is causally relevant.	F3, F4	\mathcal{RD}
(2) \overline{A} is causally relevant.	F2, F4	\mathcal{RD}
(3) \overline{B} and \overline{A} are part of the same condition.	F4, (1), (2)	\mathcal{RC}
(4) \overline{AB} and A are part of different conditions.	F4, (3), (H_2)	\mathcal{RA}

This sample proof scheme for causal inferences drawn by a sequential application of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} is to be read as follows: The ‘inference’ column lists the causal conclusions obtained in a respective step from the initial causal hypothesis and the data generated by a causal test. The ‘sanction’ column gathers the empirical data sanctioning a respective inference, while the final column lists the rules implemented within a corresponding inferential step. In analogy to this scheme, any causal inference drawn by means of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} is provable.

5.4 Limitations of the Existing Inference Procedure

The inference procedure discussed in the previous section – the 2^n -method – crucially rests on the assumption that test-factors are causally independent. The procedure presupposes the variability of test-factors in all logically possible combinations. For an analysis of n test-factors 2^n test situations are required. However, as soon as there are causal dependencies among test-factors, only a proper subset of all logically possible combinations of the test-factors may be realizable against a homogeneous background. Suppose, by means of an ff-test we are investigating the causal relevance of A and B to an effect W . The causal structure underlying these factors shall be the one depicted in figure 5.5, i.e. A is causally relevant to B which, in turn, causes C . Moreover, assume we get a test result to the effect that

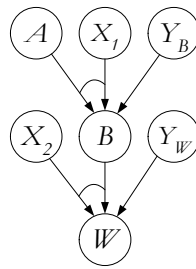


Fig. 5.5: Sample causal structure such that the causal relevance of A and B cannot be investigated by means of an ff-test.

W is instantiated in some situations, but not in all. Thus, given homogeneity, the instances of W are not accountable to any causes not containing A and B . Finally, suppose the homogeneous background of our exemplary test features a constant instantiation of X_1 . In this constellation the coincidence $A\bar{B}$ cannot be realized. For whenever AX_1 is given, there will also be an instance of B . That means, an ff-test with A and B as test-factors cannot be conducted, because not all of the 4 test situations that are necessary for an application of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} can be brought about.

The 2^n -method cannot investigate causal relevancies of arbitrary factor combinations. As we have seen in section 5.3.2, ff-tests can only be conducted on factors that operate as causes within the same layer of a causal structure. Test-factors, if they in fact are causally relevant to the effect under investigation, must either be parts of the same complex cause or of different alternative causes that are mutually causally independent. This, of course, constitutes a serious limitation to the applicability of the 2^n -method.²¹ Either it is known previously to setting up an ff-test that the envisaged test-factors actually satisfy (IND) or some trial an error procedure must be resorted to in order to determine whether all logically possible combinations of the designated test-factors are realizable. Thus, the 2^n -method either suffers from inefficiency or it presupposes a substantial insight into the causal structure under investigation, which it itself actually is designed to provide.

This latter difficulty amounts to a serious drawback for the 2^n -method. (IND) is a far-reaching *causal* assumption. Clearly, every currently known methodology regulating causal inferences operates on causal assumptions, most notably the homogeneity assumption.²² Yet, while homogeneity is an assumption about the background of an investigated causal structure, (IND) is an assumption about just that structure under investigation. It determines – prior to any tests being conducted – that there are no causal dependencies among selected test-factors. As we have seen in section 5.3.2, this amounts to an a priori exclusion of a number of possible

²¹ Ragin's (1987, 2000) QCA methodology is likewise applicable to mutually independent test-factors only.

²² Cf. e.g. Spirtes, Glymour, and Scheines (2000 (1993)), or Cartwright (1989), ch. 2.

causal structurings among test-factors. The latter are assumed to be root factors, located on the same layer – either as parts of the same complex cause or as parts of causally independent alternative causes. With a growing number of test-factors, the number of assumptively excluded causal structures increases exponentially. In case of 2 test-factors, from a total of 4 possible structurings 2 are excluded by (IND). If 3 factors are tested, (IND) excludes 24 of 29 possible structures. 4 test-factors, finally, can be grouped in 264 different ways, 258 of which are excluded by (IND). Of course, in experimental practice it happens regularly that prior causal knowledge is available to the effect that test-factors cannot be causally linked. Hence, there are experimental settings for which it can be determined that (IND) is satisfied. This, however, is not generally the case. Whenever little or nothing is known about causal processes, e.g. at the very beginning of an investigation into a certain effect's causes, (IND) amounts to a very bold assumption that presupposes much of what a causal investigation in fact should reveal.

(IND) determines all factors on which a causal test is conducted to be *directly manipulable*. No test-factor is allowed to be manipulable through some other test-factor only. This consequence of (IND) limits the applicability of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} to root factors. No doubt, the requirement concerning direct manipulability of test-factors constitutes a comprehensive limitation as regards the utility of \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} . Df- and ff-tests can be conducted on directly controllable factors only.

Finally, the 2^n -method allows for a maximum of merely *one dependent* factor. Furthermore, this single dependent factor is presumed to be identified prior to conducting causal tests.²³ Df- and ff-tests can only be set up if one factor of a set of coincident factors is specified to be the effect. Even though this may often conform to experimental practice, it by no means generally must be the case. It may well happen that all data available prior to a causal investigation consists in sets of coincidentally instantiated factors such that none of them is tagged to be the effect.

All in all, these are significant limitations. The ultimate reason for the introduction of (IND) in section 5.3.2 has been the ambiguity of causal inferences drawn from ff-test tables without (IND). However, we have seen that these ambiguities do not result from *mt*-equivalencies of inferred minimal theories.²⁴ Rather, it may happen that the differences between inferred minimal theories do not show up against the background of only four test situations selected for an ff-test. Reconsider the two minimal theories (5.1) and (5.2) compatible with the coincidence data gathered in table (a) of 5.4 on page 205 above. (5.1) and (5.2) are not *mt*-equivalent, nonetheless their empirical differences do not appear in the four selected test situations. However, if more data were collected, sooner or later the extended coincidence data would cease to be compatible with both (5.1) and (5.2).

²³ Again, the same holds for Ragin's (1987, 2000) QCA.

²⁴ Cf. page 206.

Moreover, the previous chapter has shown that even if ambiguities are not resolvable within limited factor frames, extensions of the latter will decide among competing causal hypotheses. Therefore, introducing additional causal assumptions besides homogeneity is not a precondition of a causal inference procedure that unambiguously identifies causal structures. Instead of resorting to a far-reaching causal assumption as (IND), the ambiguities with respect to assigning a causal structure to an ff-test table as depicted in 5.4 can also be resolved by generating further empirical data. This non-causal strategy to resolve ambiguities will be followed by the inference procedure to be proposed below.

5.5 Coincidence Analysis

5.5.1 The Basic Idea

The new inference procedure – *coincidence analysis* (CA) – to be developed in this section essentially differs from the 2^n -method as regards its abandonment of (IND). As an immediate consequence thereof, CA does not rely on any assumption about the causal structure under investigation. None of the possible structurings of the test-factors are excluded prior to analyzing the empirical data. Homogeneity of test situations then amounts to the only causal assumption entering CA – and homogeneity concerns the *background* of the investigated structure not the latter itself. Test-factors are neither required to be causally independent, nor to be root factors or directly manipulable. Furthermore, it will not even be presupposed that a particular factor within an analyzed factor frame is specified to be the effect. (IND) will be replaced by a requirement to collect *all available* empirical data about an investigated causal process. Ambiguities in light of limited coincidence data shall thus not be resolved by means of an additional causal assumption, but by simply requiring further data. Of course, whenever n test-factors are antecedently known to be causally independent, the data collection can be confined to 2^n test situations as required by the 2^n -method. In this sense, CA can be seen as a generalization of the 2^n -method. Yet, contrary to the latter, CA is applicable when nothing at all is known about a causal process. It allows the coincidences under investigation to be the result of structures consisting of an arbitrary amount of layers.

CA takes up the results obtained at the end of chapter 4. It assigns minimal theories to sets of coincidences that are empirically realizable by systematically varying the factors within the factor frame constituting the structure under investigation. CA can essentially be seen as a three-step procedure:

1. All empirically realizable coincidences within the factor frame of the investigated causal structure are collected.
2. One (or more) minimal theory(ies) is/are assigned to that set of coincidences.
3. The minimal theories obtained in step 2 are causally interpreted such that disjuncts in their antecedent are interpreted as (alternative) complex causes of the factor in the consequent.

CA operates on the same data as the 2^n -method: coincidences of the factors involved in a causal process whose structure is to be revealed. Contrary to the 2^n -method, however, the data fed into CA is not required to mark one factor as the effect. Based on its input data, CA simply determines for each factor Z in the analyzed frame which dependencies hold between Z and the other factors in the frame. Most of these dependencies will turn out not to be causally interpretable. The possibly causally interpretable dependencies are subsequently minimalized and expressed in terms of minimal theories, which, finally, are straightforwardly causally interpretable as shown in chapter 3. The intricate part of CA is the assignment of minimal theories to sets of coincidences. That is not to say that e.g. step 1, the data collection, is unproblematic. After all, it can only be guaranteed that CA generates an unambiguous output, provided that all empirically realizable coincidences of the factors involved in an analyzed causal process are in fact collected. That, of course, may be seriously hampered by practical constraints as experimental or observational limitations. However, data collection is not a matter of causal reasoning. Relative to incomplete empirical data, no unambiguous causal inferences are possible. Nonetheless, CA is able to assign sets of causal structures to incomplete coincidence data such that each structure in such a set is compatible with the coincidences constituting the input data.²⁵ That means, even if there only is a guarantee for unambiguous outputs relative to complete input data, CA is able to at least select a set of causal structures that all could possibly generate the incomplete data.

Since step 3 has been the topic of chapter 3, nothing much about this final step of CA will be said in the following. Basically, expressing the dependencies holding among the factors in an analyzed frame in terms of minimal theories is merely a syntactical convention. Other equivalent syntactical forms would be available. The double-conditional form is preferred simply because there exists a straightforward causal interpretation rule for that form: The antecedent of a double-conditional is constituted by the alternative (complex) causes of the effect mentioned in the consequent.

5.5.2 Input Data

The data CA operates on is of the same sort as the data that enters the 2^n -method: records of coincidentally instantiated factors in causally homogeneous²⁶ test situations, or coincidences for short. While the coincidence information on which \mathcal{RD} , \mathcal{RC} , and \mathcal{RA} are brought to bear is displayed by means of ff-test tables as table (a) of 5.5, the input data is simply listed in case of CA, as illustrated by table (b) of 5.5. Ff-test tables only explicitly mention the test-factors, while occurrences of the effect – C in the exemplary case (a) – are marked by ‘1’ and ‘0’ in the fields of the

²⁵ Cf. section 5.8.

²⁶ The notion of causal homogeneity will be adapted for the purposes of CA below (cf. section 5.5.3, p. 231).

V_6	B	\overline{B}
A	1	1
\overline{A}	1	0

(a)

A	B	C
1	1	1
1	0	1
0	1	1
0	0	0

(b)

Tab. 5.5: Table (a) collects the coincidence information as required by the 2^n -method, table (b), in contrast, lists the same information in terms of CA.

table. The mere form of this tabular exposition of the data induces a distinction between possible causes and effects. Test tables representing test situations in a way that suits the 2^n -method presuppose the variability of test-factors in all logically possible combinations, i.e. the independence of test-factors. As mentioned above, CA does neither presuppose the independence of any factors nor the specification of one factor as effect. For this reason, the simple listing as illustrated in table (b) is chosen for the exposition of data to be processed by CA. Tables as (b) will be labelled *coincidence lists*. The rows in a coincidence list shall be numbered starting with the first row below the title row. The row constituted by “1 1 1” in (b) is row 1 (R1), the row featuring “1 0 1” is row 2 (R2), and so on. Contrary to a table of kind (a), a ‘1’ in a coincidence list does not symbolize the occurrence of effects not explicitly referred to in the list. Rather, in table (b) a ‘1’ in the column of, say, factor A represents an instance of A , a ‘0’ in that same column correspondingly symbolizes the absence of such an instance. Columns of coincidence lists thus record instances and absences of the factor mentioned in the title row, while the rows following the title row specify coincidences of the factors in the title row. For example, the first row R1 of (b) records the coincidence ABC , the following row R2 indicates the coincidence $A\overline{B}C$. In contrast to ff-test tables, the syntax of coincidence lists does not have any implicit implications. All factors are treated alike, neither systematic variability nor a cause-effect distinction is presupposed by these lists. Nonetheless, both tables, (a) and (b), display the exact same coincidence information. There is a one-to-one correspondence between the fields of (a) and the rows of (b). Field 1 corresponds to row 1, field 2 to row 2, and so on. However, while for every data table fed into the 2^n -method there is an equivalent coincidence list, not every coincidence list is expressible by means of data tables as required by the 2^n -method. For instance, a coincidence list (b’) that results from (b) by removing the latter’s last row still constitutes a well-formed coincidence list, yet there does not exist a corresponding well-formed ff-test table. Ff-test tables are only well-formed if all logically possible variations of the test-factors are empirically realized in test situations.

As mentioned above, coincidence lists record coincident instantiations of the factors in their title rows. Accordingly, R1, for instance, can be read as “There is

a coincident instantiation of the factors A , B , and C ”; R2 correspondingly states “There is a coincident instantiation of A and C while there is no instance of B ”. Formally put, where K represents the coincidence relation, R1 and R2 state:

$$\exists x, y, z(Ax \wedge By \wedge Cz \wedge Kxyz) \quad (5.4)$$

$$\exists x, z \neg \exists y(Ax \wedge By \wedge Cz \wedge Kxyz) \quad (5.5)$$

As we have seen in section 2.3.4, the notion of a coincidence is of crucial importance both for probabilistic accounts of causation and for MT. Despite its importance, the problems this notion raises are hardly ever tackled in theories of causation. Most of all it is unclear how “...is coincident with...” is to be interpreted such that for any two events it is determinable whether they are coincident or not. Interpretations of the coincidence relation are thus problematic, because this relation is presupposed by conceptual analyses of causation, while, at the same time, it seems difficult to spell out “...is coincident with...” in *non-causal* terms. Section 3.3 has proposed a treatment of the coincidence relation K on a par with any causally relevant factor – just that ordinary causal factors correspond to unary predicates, while K is n -nary, where n is the number of coincidentally instantiated factors. Yet, apart from this difference, the coincidence relation shall be taken to be a normal non-redundant part of a minimally sufficient condition contained in a minimal theory. This treatment of K has the important advantage that it dispenses with having to provide a fixed interpretation for K . Depending on the factors analyzed, coincidence can thus be interpreted flexibly. If factors with microscopic instances are causally analyzed, events will be termed “coincident” if they happen within a very small spatiotemporal interval, while in case of macroscopic factors, coincidence is to be understood in terms of a more extended interval.

Endorsing this treatment of coincidence when it comes to collecting empirical data in the course of causal reasoning, however, poses a serious problem. Minimal theories are to be inferred from coincidence lists. Thus, whether a factor – including K – in fact is a non-redundant part of a minimally sufficient condition contained in a minimal theory, is to be inferred from coincidence lists. That means, if coincidence is understood in terms of an ordinary non-redundant part of minimally sufficient conditions, it is determinable what interpretation has to be afforded to K only after having assigned a minimal theory to a coincidence list. However, collecting the data fed into CA already presupposes clarity on the way coincidence is to be understood relative to the analyzed factor frame, for, as exhibited by (5.4) and (5.5), filling in the rows of coincidence lists amounts to claiming that respective factors are *coincidentally* instantiated. The input data of CA consists of nothing but *coincidentally* instantiated factors. Hence, dispensing with explicating coincidence and with fixing an interpretation of that relation by merely treating it on a par with ordinary unary factors leads into a sort of epistemological circle when it comes to the compilation of coincidence lists. The compilation of these lists presupposes what is only inferrable from these lists.

This problem is yet accentuated by the fact that minimal theories not only require the factors within a complex cause to be coincidentally instantiated, but moreover they assess that instances of the causes are related in terms of relation R to the instances of the effect.²⁷ Of course, minimal theories are only assignable to coincidence lists if the events constituting these coincidences are not only related in terms of K , but also in terms of R . R , however, could not be treated on a par with unary factors in chapter 3, for R appears in the consequent of a minimal theory. Its interpretation had to be fixed to the admittedly vague "...occurs within the same spatiotemporal frame as ...". The events assembled in coincidence lists must be assumed both to be coincidentally instantiated and to occur within the same spatiotemporal frame. Therefore, not only the epistemological circle with respect to the interpretation of coincidence, but also the determinately fixed interpretation of R seem to induce that the relation holding among the instances of the factors assembled in a coincidence list cannot be left uninterpreted.

At the same time, as mentioned above, in experimental practice the understanding of coincidence appears to vary with the sort of causal process under investigation. Thus, if it should turn out that the coincidence relation needs to be given a fixed interpretation in order to be able to collect causally analyzable data and that such an interpretation – in some sense or another – rests on the causal process under investigation, we are facing a serious danger to the whole enterprise of analyzing the notion of causation. For if K is to be understood relative to a respective causal process, the interpretation of K presupposes clarity on the notion of a causal process and thus of causal relevance. Clearly, if an understanding of K , which essentially appears in minimal theories – the core of our analyses of causal relevance – and which is a necessary precondition of the collection of causally analyzable data, itself presupposes clarity on the notion of causal relevance, the theory of causation presented in the present study becomes fundamentally circular. Obviously, this fatal consequence to the endeavor undertaken here must, by all means, be prevented.

Several solutions to the problem at hand can be thought of. For instance, it might be claimed that, depending on the factor frame under investigation, non-causal interpretations of K must just be stipulated. Say, for microscopic factors "... is coincident with ..." is to be understood along the lines of "... is spatiotemporally contiguous to ...". Then a function could be defined that assigns increasing spatiotemporal intervals to K proportionally to increasing size of the instances of analyzed factors. Yet, such a stipulation, apart from being utterly arbitrary, does not yield interpretations of K that suit all causal processes. Certain macroscopic factors as – classically – moving billiard balls call for contiguity of their instances in order for corresponding effects to occur, others as wet streets and dysfunctional traffic lights cause accidents even when their instances are located several meters apart. Notwithstanding the fact that microscopic factor frames usually call for

²⁷ For details on R see section 3.2.

small spatiotemporal intervals with respect to K , while macroscopic frames tend to allow for more extended intervals, this is by no means generally the case.

As a consequence thereof, it might be proposed that examples of the traffic lights type call for further specifications. It could be held that the combination of dysfunctional traffic lights and wet streets cannot be seen as a direct cause of car crashes, rather, a lot of intermediary factors operate on causal paths between these coarse-grained factors. For instance, dysfunctional traffic lights cause absences of light signals travelling from above the intersection to the eye of the driver, who thence hits the brakes too late such that, when he finally does, the wheels slide across the wet street causing the crashing into a crossing car. If this specification is carried out far enough, thus the second proposal as to explicating K might continue, a causal chain of contiguously instantiated factors will result. Hence, the interpretation of "...is coincident with..." could generally be fixed to "...is spatiotemporally contiguous to...". However, this way of fixing the interpretation of K on non-causal grounds would imperatively induce to generally analyze causal structures on highly fine-grained levels of specification.²⁸ Causally analyzable data could only consist of contiguously instantiated factors, other factor frames would not be causally analyzable to begin with. The level of specification of causal analyses would not be open and adjustable to the interests and requirements of a given investigation, but would a priori be fixed to high specificity. Furthermore, it is far from clear whether such contiguous sequences of events can be found in case of all complex causes. Consider, for instance, the elliptical motion of the planets caused by complex causes consisting of the gravitational forces of the planets and the corresponding forces of the sun, or the budget deficit of a firm in combination with high tax rates which cause the firm to lay off a certain number of its employees. In what sense can the instances of these complex causes be seen as connected by sequences of contiguous events? Spatiotemporal contiguity of events, thus, is not a necessary condition for corresponding factors to become causally effective as a complex cause.

Stipulatively fixing K to a specific spatiotemporal interval is not only arbitrary, but moreover inadequate. Depending on the context, "...is coincident with..." is to be understood along the lines of "...is spatiotemporally contiguous to..." or some non-zero spatiotemporal interval may well or even must be admissible. In order to avoid having to fix K to some arbitrary interval, it might be argued that in scientific practice it usually is sufficiently clear when events are spatiotemporally thus related that they could possibly causally interact, even though a clear-cut analysis of that relation cannot be provided. Essentially, that would amount to how most presently known accounts of causation handle the coincidence relation: They – in one way or another – presuppose coincidence as an unanalyzed primitive relation, that, according to a corresponding context, is somehow interpretable on

²⁸ Similar consequences arose in case of fixing of the interpretation of R to spatiotemporal contiguity – as has been proposed as a solution to the chain-problem in section 4.6.2, p. 178.

non-causal grounds.²⁹ All that is usually provided is an upper bound of the interval admissible for "... is coincident with ...": According to the theory of relativity, one of two causally interacting events must occur in the past light-cone of the other ('no action at a distance').

This pragmatic way of handling the interpretation problem with respect to K faces two difficulties: First, even though interpretations of K may often be sufficiently clear in scientific practice, this by no means generally must be the case; and second, especially in cases when determining an interval for "... is coincident with ..." is unproblematic, it cannot be excluded that this manifest interpretation is induced by prior causal knowledge about the process under investigation. For instance, consider again the striking of a match, factor A , and the presence of oxygen, B , as a complex cause of a match catching fire, C . Whether A and B in fact constitute a complex cause of C would certainly not be tested by looking at instances of A in Bern and instances of B in London, not even instances of A in one room and instances of B in a neighboring room would be of interest. It is plain in this case that A and B must be contiguously instantiated in order for them to be causally effective. Nobody would be ready to deny AB the status of a complex cause for C if he would be presented with an instance of A in a vacuum chamber and a simultaneous instance of B only millimeters away without the match catching fire. AB would, notwithstanding this test result, still be seen as a complex cause of C , simply because the two causal factors are not coincidentally instantiated in the course of the purported test situation. The reason for the incontrovertibility of having to understand K in terms of spatiotemporal contiguity in this case lies in the fact that the causal processes taking place during combustion are sufficiently well known. Due to this prior causal knowledge, it can be excluded that striking a match and the presence of oxygen could possibly interact through the glass of a vacuum chamber. This is a clear case of prior causal knowledge inducing an interpretation of K .

However, as we have seen above, if it should turn out that the notion of coincidence presupposes clarity on causal relevance, a reductive analysis of causation in terms of regularities or probabilistic correlations will be impossible, for it will be doomed to circularity. As long as prior causal knowledge is only implemented to interpret K as a heuristic means facilitating such an interpretation in obvious cases, as the match example above, no difficulties arise. Yet, an analysis of causal relevance on regularity theoretic grounds can only succeed, given that, at least in principle, there is a way to determine spatiotemporal intervals for K *without* recourse to causation. All in all, this shows that the problem of explicating "... is coincident with ..." cannot be solved in a way that meets the purposes of a regularity theory by stipulating an interpretation of K or by simply assuming sufficient

²⁹ For example, in the course of collecting statistical data for probabilistic causal analyses, it is generally assumed that the instances of the factors in test as well as in control groups are coincident, cf. Spirtes, Glymour, and Scheines (2000 (1993)).

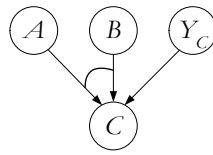


Fig. 5.6: Exemplary causal graph. For simplicity's sake, it is assumed that AB constitutes a complete cause of C .

clarity on such an interpretation in practical contexts, as long as it cannot be guaranteed that explicating "... is coincident with..." does *not essentially* presuppose an analysis of causal relevance.

Such a method to determine a spatiotemporal interval for K on non-causal grounds shall now be proposed. In order to do so, let us first look at a simplifying example. Suppose, we are investigating the causal structure regulating the behavior of three factors A , B , and C . The structure we are looking for shall be the one depicted in figure 5.6. Thus, AB constitutes a complete complex cause of C , which, apart from AB , has a number of unknown alternative causes that are not going to be investigated. We know that constituents of a complex cause must be instantiated within some spatiotemporal interval K in order to become causally effective and that causes and their effect occur within the same spatiotemporal frame (R). However, assume we have no prior causal knowledge that could help determining spatiotemporal intervals for K and R relative to the factor frame considered here. How are we to collect causally analyzable coincidences of these three factors without knowing what spatiotemporal relations have to subsist among their instances? Such as not to invoke any causal intuitions that would help determining spatiotemporal intervals fitting the factor frame under consideration, I shall not provide any interpretations for A , B , and C . All we can be sure of is that instances of the test-factors must be collected such that locality, i.e. the principle of no action at a distance, is satisfied. Suppose, we randomly choose some interval for K , say "... occurs within 1 hour and 2 meters of ...", and for R , say "... occurs within 2 hour and 4 meters of ...". Let us presume that both of these intervals are far too wide for any causal dependencies among A , B , and C to crop up. How does such an inadequate interpretation of K and R affect our data collection?

Before this question can be answered, one assumption on which **CA** will be shown to rest in section 5.5.3 must be anticipated at this point, for it will prove to be of crucial importance to the discussion below. It will be assumed that **CA** faces no practical limitations as to data collection. That means that all coincidences of test-factors that are compatible with the causal structure regulating their behavior are in fact observable. In this context, we shall speak of the *principle of empirical exhaustiveness*, or (PEX) for short.³⁰

³⁰ For a detailed discussion and justification of (PEX) cf. section 5.5.3, p. 228 et seq.

(PEX) induces an answer to the question raised above along the following lines: Non-coincident instantiations of A and B are not sufficient for C , which means that we will find or be able to realize instances of AB without a corresponding instance of C . Instances of $AB \vee Y_C$ that are not related in terms of R to C are not necessary for C , which is to say that we shall find or be able to realize instances of C without a corresponding instance of $AB \vee Y_C$. Apart from these combinations, there will also be co-occurrences and absences of all three factors within the spatiotemporal intervals mentioned above. In short, the result of our inadequate interpretation of K and R will be that we find or are able to realize co-occurrences of A , B , and C , that conform to the above spatiotemporal intervals in *all logically possible* combinations. This finding, of course, essentially hinges on (PEX). For in reality it happens every now and then that data collection faces a host of practical limitations to the effect that not all empirically possible coincidences over an investigated factor frame are in fact observable. However, practical limitations of data collection are not an issue of causal reasoning, and besides, practical limitations with respect to data collection are likely to be identifiable as such. Based on inexhaustive data no reliable causal inferences can be drawn in principle, hence, we shall neglect such practical limitations in what follows.

Thus, (PEX) ensures that no dependencies among the test-factors can emerge relative to overly wide spatiotemporal intervals for K and R . The factors in our investigated exemplary frame, therefore, appear to be independent. That means that our data collection yields a *complete coincidence list* over the factor frame consisting of A , B , and C , as illustrated in table (a) in 5.6.

Complete coincidence list: A coincidence list over a factor frame of n factors is called *complete* iff it contains all logically possible combinations of the involved factors, i.e. iff it is constituted by 2^n rows.

A	B	C	A	B	C
1	1	1	1	1	1
1	1	0	1	1	0
0	1	1	0	1	1
1	0	1	1	0	1
1	0	0	1	0	0
0	1	0	0	1	0
0	0	1	0	0	1
0	0	0	0	0	0
(a)			(b)		

Tab. 5.6: Table (a) is a complete coincidence list over the factor frame of the graph in figure 5.6. It is the result of an inadequate interpretation of K and R . As soon as K and R are adequately interpreted, the structure in 5.6 generates table (b).

The same will result from any inadequate interpretation of K and R . Whenever the spatiotemporal intervals chosen for K and R are too wide (or too narrow) for causal dependencies to emerge, i.e. unsuited for factors of complex causes to interact and for events to be related as causes and effects, all logically possible combinations of the investigated test-factors are empirically realizable.

However, as soon as K and R are thus adjusted that factors constituting a complex cause can possibly interact and that events can possibly be related as causes and effects, not all logically possible combinations of A , B , and C can be found or realized any more. Dependencies begin to emerge. When K is assigned a spatiotemporal interval such that instances of A and B in fact can interact and R is understood in terms of an interval such that instances of A , B , and C can actually be related as causes and effect, we shall no longer find instances of AB without an instance of C , for the structure in figure 5.6 determines a coincident instantiation of AB to be sufficient for C . Therefore, relative to a suitable interpretation of K and R , the combination $AB\bar{C}$, i.e. row 2 of table (a), is not empirically realizable any longer. If K and R are properly interpreted, our data collection cannot yield a complete list over the factor frame at hand any more.

This finding suggests the following method of determining adequate spatiotemporal intervals for K and R : When collecting causally analyzable coincidence data, interpretations of K and R are to be systematically varied until dependencies appear, i.e. until data collection does not render complete lists over investigated factor frames. Without any prior causal knowledge there is no alternative to this kind of trial-and-error methodology for pinning down intervals for K and R . As any mere trial-and-error methodology, interpreting K and R is a highly involved task, which faces severe practical limitations. There is an infinite number of spatiotemporal intervals. Thus, it is impossible to test out all feasible interpretation candidates. At some point of repeated unsuccessful attempts to fine-tune K and R in a way that reveals dependencies, the trial-and-error procedure just has to be given up. There is no guarantee that a continued variation of spatiotemporal relationships between instances of analyzed factors ever reveals dependencies. Two reasons could be accountable for such a negative result: (1) there in fact are no causal dependencies among the factors under investigation, or (2) due to practical limitations, the attempt to suitably interpret K and R has to be suspended. In section 5.3.1 (p. 201) we have already encountered a first reason as to why a procedure of causal reasoning that operates on coincidence data alone cannot infer causal irrelevancies. At this point we find a second reason for this inferential limitation. The fact that certain coincidence data does not reveal causal dependencies can be due to either (1) or (2) and it is not determinable which of the two possible reasons in fact is to be held responsible for complete coincidence lists. Therefore, the procedure of causal reasoning to be developed in the following has to abstain from inferring anything with respect to complete coincidence lists.

I hold that this limitation as regards the causal analyzability of coincidence information affects all procedures of causal reasoning that process coincidence data.

There cannot be such a thing as a *complete* causal inference procedure based on coincidence information only, i.e. a procedure such that if the behavior of the factors in a given frame \mathcal{F} is in fact regulated by a causal structure, the procedure infers that structure – in a finite number of steps – from coincidence data over \mathcal{F} . A procedure inferring causal structures from coincidence information can only claim *correctness* given the validity of its assumptions: If the procedure assigns a causal structure to the factors in frame \mathcal{F} , the behavior of the factors in \mathcal{F} in fact is regulated by that inferred structure provided the assumptions that enter the procedure – e.g. homogeneity – are warranted. There is no guarantee that nature provides us with suitable coincidence information for all existent causal structures. In light of the empirical nature of the input data, the applicability of our inference procedure has to be confined to the availability of suitable coincidence information. As long as the factors in an analyzed frame are found to be instantiatable in all logically possible combinations, no dependencies holding among these factors can be diagnosed. This, however, does not mean that these factors are causally independent, rather, it merely suggests that the available data is unsuited for a causal inference.

Clearly, there might be useful heuristics that regulate the systematic variation of K and R in a way that allows for reliably estimating whether the inability to generate incomplete coincidence lists stems from (1) or (2). Such heuristics, however, shall not be discussed in the present context. For the purposes of the study at hand, it suffices to note that K and R can be assigned spatiotemporal intervals on non-causal grounds – at least in principle. Interpreting K and R , thus, does not essentially presuppose clarity on the notion of causal relevance. We shall not claim more than correctness for **CA**. Most importantly, hence, unsuited interpretations of K and R are harmless to **CA**, for they *do not induce wrong causal inferences*. Collecting coincidence data based on inadequate spatiotemporal intervals attributed to K and R yields complete coincidence lists, which, in turn, do not reveal any dependencies among the involved factors. No minimal theories are assignable to complete coincidence lists.

Incomplete lists only arise from suitable intervals attributed to K and R . Apart from demonstrating that **CA** in fact does not infer anything from complete lists, we shall therefore not be concerned with complete lists in the following any more. The input data of **CA** is constituted by incomplete coincidence lists, which is to say that suitable interpretations of K and R are available. That the events recorded in the rows of incomplete lists are thus related that a possible causal interaction among them is not excluded merely due to spatiotemporal indispositions, is ensured by the mere fact that the corresponding list is incomplete and by (PEX). The events appearing on one and the same row in a coincidence list are coincident (relative to the corresponding factor frame) and related in terms of R . All relational constraints imposed by minimal theories are hence assumed to be satisfied by events recorded in coincidence data to be processed by **CA**. These relational constraints are necessary preconditions for a causal interaction among events, but they are not sufficient for such an interaction. Events only causally interact provided that the

corresponding factors are related in terms of causal relevance – and such a relationship, as we have seen in chapter 3, is neither induced by K nor by R , but only by the membership of these factors in minimal theories.

5.5.3 Presuppositions

As the 2^n -method, **CA** rests on a number of important presuppositions. First of all, **CA** assumes the causal relation to be deterministic as expressed by the Principle of Determinism (PD) and the Principle of Causality (PC). Likewise, the Principles of Relevance and Persistent Relevance are presumed to hold. Since these principles have been extensively discussed in section 2.4, they can be taken for granted at this point. Secondly, homogeneity of test situations has to be assumed by **CA** as well. In section 5.2 we have only defined homogeneity for four test situations, yet as indicated above, **CA** processes far more than four test situations. Therefore, homogeneity will be amended to suit the purposes of **CA**. Furthermore, it has already been advanced in section 5.4 that one of the crucial assumptions backing the 2^n -method is given up by **CA**. Due to its essentially causal and far-reaching nature, the assumption concerning causal independence of test-factors (IND)³¹ is discarded. In its place, as we have seen in the previous section, **CA** shall be backed by explicitly assuming that data collection does not face any practical limitations (PEX).

Exhaustiveness of Coincidence Data

In fact, any procedure of causal reasoning, in some way or another, assumes that its input data is exhaustive as required by the respective procedure. Probabilistic procedures presume the availability of probability distributions over all exogenous variables, or the 2^n -method relies on the realizability of all 2^n test situations. Nonetheless, assumptions as regards the exhaustiveness of empirical data are hardly ever made explicit in studies on causal reasoning.³² Such an implicit taking for granted of the suitability of input data, however, will not do for the present context, for, as the previous section has shown, the absence of practical limitations as to data collection is of crucial importance to the theoretical fundament of **CA**.

Two variants of (PEX) can be distinguished: a strong and a weak one. The strong variant corresponds to what has been anticipated in the previous section. We shall mainly presuppose this strong version of (PEX).

Strong Principle of Empirical Exhaustiveness (PEX): The collection of empirical data to be processed by **CA** faces no practical limitations whatsoever. That means, within the homogeneous background of a causal test, all coincidences of test-factors that are compatible with the causal structure regulating their behavior are in fact observed.

³¹ Cf. section 5.3.2, p. 205.

³² One exception is Ragin (1987, 2000). He discusses at length how limited empirical data affects causal reasoning.

The main reason for the introduction of (PEX) lies in its central role in the process of non-causally interpreting K and R . Incomplete coincidence lists shall be taken to indicate that spatiotemporal intervals for K and R are properly fine-tuned. This, obviously, would not be possible, if incomplete lists might also result from practical limitations with respect to data collection. In the following, we shall see that (PEX) also serves the unambiguity of causal inferences. Minimal theories are only unambiguously assignable to coincidence lists provided that the latter are assumed to be empirically exhaustive in the sense expressed by (PEX). However, while (PEX) is a necessary condition for drawing *unambiguous* inferences, it is not a necessary condition for drawing (ambiguous) causal inferences from coincidence lists. If inexhaustive lists are processed by CA, as will be shown in section 5.8 below, more than one minimal theory will be assigned to such a list. The number of minimal theories assigned to an inexhaustive list depends on the logical possibilities of complementing a respective inexhaustive list in a causally interpretable manner. Thus, while it is impossible to infer a single causal structure from an inexhaustive coincidence list, a class of structures can be inferred such that all of its members are compatible with the coincidences recorded in the inexhaustive list. Assigning classes of causal structures to inexhaustive lists, of course, also is a form of causal inference. Such inferences might prove to be of great use in scientific practice, for they at least shed light on what structures *cannot* underly an investigated factor frame. Therefore, CA does not necessarily have to be based on (PEX). The exhaustiveness principle might be weakened such that its validity is limited to the process of finding suitable interpretations of K and R .

Weak Principle of Empirical Exhaustiveness (PEX_w): The collection of empirical data serving the interpretation of K and R faces no practical limitations whatsoever. That means, within the homogeneous background of a causal test all coincidences of test-factors that are compatible with the causal structure regulating their behavior are in fact observed in the course of fine-tuning spatiotemporal intervals for K and R .

Or put differently, when it comes to interpreting K and R , incomplete coincidence lists may only result from suitable intervals chosen for these two relations, not from practical limitations with respect to data collection. Yet, as soon as interpretations of K and R are available, empirical exhaustiveness does not need to be demanded any longer. Apart from interpreting K and R , (ambiguous) causal reasoning is possible even based on inexhaustive data. As already mentioned above, however, we will mostly be interested in unambiguity in the following. For this reason, (PEX) will be used in its strong variant.

No doubt, (PEX) is often violated in scientific practice. Exhaustive data collection may fail for a host of different reasons. Financial or technical resources may happen to be limited in experimental sciences or nature may be found not to provide sufficient data in non-experimental sciences. Insufficient data is likely to be one of the main reasons for hampered causal interpretability of that data. Relative to insufficient data, unambiguous and reliable causal reasoning simply is

impossible. Data collection, however, is not part of causal reasoning, but rather a precondition thereof. That is why (PEX) is endorsed in the present context, which is concerned with matters of causal reasoning only. Clearly, (PEX) constitutes a sweeping idealization with respect to data collection. Such an idealization, however, may prove to be useful in many practical contexts. It can be implemented as a gauge by means of which concrete data collections can be measured and thus evaluated. Section 5.8 below will be concerned with causal reasoning based on inexhaustive empirical data.

Homogeneity

As coincidences analyzed by the 2^n -method, the input data of CA is assumed to be causally homogeneous. Within inhomogeneous backgrounds, i.e. with so-called *confounders* arbitrarily meandering between presence and absence, a causal interpretation of test situations is impossible – as demonstrated in section 5.2. A confounder is a causally relevant factor that is not contained in the analyzed factor frame and whose uncontrolled variation may invoke causal fallacies, i.e. its variation may lead to a fallacious attribution of causal relevance to test-factors.³³ We have seen that the background of a four-field test can either be confounded by hidden alternative causes of the investigated effect or by uncontrolled co-factors of test-factors.³⁴ The set of factors that must be allowed to vary, in contrast, includes the test-factors, the investigated effect, genuine test-factor causes and intermediate factors located on a causal chain between test-factors and the effect. CA can implement the exact same distinction between confounders and factors that are allowed to vary as does the 2^n -method. Both procedures implement analogous methodologies – systematic variation of test-factors – in order to achieve the same goal: determining, first, whether test-factors are causally relevant and, if that is the case, second, where in a causal structure these causally relevant test-factors are located.

Despite this common ground, there is a difference between the two procedures that affects the notion of causal homogeneity implementable by the 2^n -method and CA, respectively. The above considerations show that what is to count as a confounder depends on what factor is investigated as effect. Thus, the notion of causal homogeneity is dependent on the identity of a corresponding effect. Accordingly, (HC*) is defined *for* a given effect.³⁵ However, I have already indicated that one of the salient differences between the 2^n -method and CA consists in the latter's indeterminacy with respect to what factors in an analyzed frame are to be seen as operating as effects and causes within the underlying structure. Contrary to the 2^n -method, CA identifies dependencies subsisting between *all* factors in the analyzed frame. Some of these dependencies are causally interpretable, while most of them will turn out not to be thus interpretable. To this end, CA conducts several runs through an analyzed coincidence list, each time assuming another factor to be the

³³ For more about the notion of a confounder see Baumgartner and Graßhoff (2004), pp. 234-238.

³⁴ On the notion of a *co-factor* see section 5.3.2.

³⁵ Cf. p. 203 above.

ultimate effect of the underlying structure. Some of the factors in a given frame can be excluded from possibly operating as effects on a priori grounds, because they are universally instantiated or absent – which precludes them from causal analyzability due to a violation of the Principle of Relevance (PR). Details of the identification of potential effects will be discussed in section 5.5.4 below. For now, all that matters is that CA does not presuppose a single factor to be the effect, but, rather, specifies a whole set \mathbf{W} of factors that possibly operate as effects within the causal structure under investigation. Causal homogeneity thus needs to be defined for each factor in \mathbf{W} . Apart from this deviation from (HC*), causal homogeneity suiting the needs of CA must be kept open as regards the number of test situations, for, contrary to ff-tests, CA is not confined to analyze a maximum of four test situations. In principle, the number of test situations processable by CA is open to any finite integer.

Homogeneity (HC): The background of m test situations generating a causally analyzed set of coincidences over a factor frame consisting of n factors Z_1, Z_2, \dots, Z_n and containing the set \mathbf{W} of potential effects is causally homogeneous iff for each $Z_i \in \mathbf{W}$ and the set $\mathbf{T} = \{Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n\}$ of test-factors the following two conditions hold:

- (1) of all *minimally sufficient conditions* X_k for Z_i (or $\overline{Z_i}$) at least one conjunct is absent in one of the test situations iff at least one conjunct of X_k is absent in all other $m - 1$ test situations as well, where X_k satisfies the following conditions:
 1. none of Z_1, Z_2, \dots, Z_n and $\overline{Z_1}, \overline{Z_2}, \dots, \overline{Z_n}$ is part of X_k ,
 2. no part of X_k is a genuine test-factor cause or an intermediate factor between a test-factor and Z_i ,
 3. the parts of X_k are *causally relevant* to Z_i (or $\overline{Z_i}$).
- (2) the rest of a complex cause X_l of Z_i (or $\overline{Z_i}$) such that $Z_1, \dots, Z_j \in \mathbf{T}$, $j \geq 1$, are part of X_l , does *not vary* in test situations featuring at least one of Z_1, \dots, Z_j .

More concisely put: A coincidence list over a factor frame consisting of Z_1, Z_2, \dots, Z_n , which is analyzed by CA, is assumed to be homogeneous with respect to possible confounders not contained in Z_1, Z_2, \dots, Z_n . Depending on the specific $Z_i \in \mathbf{W}$ analyzed in the course of a given run of CA, different factors are to be seen as confounders and, accordingly, must be homogenized. Generally: Input data processed by CA is assumed to be the result of homogeneous test situations in the sense of (HC).

(HC) excludes a number of coincidence lists from causal analyzability. Contrary to ff-test tables processed by the 2^n -method, the coincidence lists fed into CA may well reveal the background of test situations to be causally inhomogeneous. Consider, for instance, the tables in 5.7. Assume B to be an effect of the causal structure generating list (a) in table 5.7. A comparison of the test situations

A	B	C	A	B	C
1	1	1	1	0	0
1	0	1	0	1	0
0	1	1	0	0	1
1	1	0	0	0	0
(a)			(b)		

Tab. 5.7: Two coincidence lists that cannot be causally analyzed, for none of the involved factors can be interpreted as effect of an underlying causal structure such that (HC) is satisfied.

recorded in row 1 and 2 of that list shows that, if B in fact were the effect of the underlying structure, the test situations recorded in (a) would violate (HC). The only factor varying in R1 and R2 is B ; no other factor in the frame $\{A, B, C\}$ is accountable for that variation of B , therefore it must be due to a varying alternative cause of B in the unknown background of list (a). That means, assuming B to be an effect contradicts the homogeneity assumption. If B is taken to be a root factor of the underlying structure, (HC) is not violated. Thus, assuming (HC) to hold for list (a) implies that B cannot be seen as a possible effect. The same holds for the other two factors in $\{A, B, C\}$. In R1 and R3 A is the only varying factor, while no other factor, apart from C , varies in R1 and R4. Hence, there is no factor in list (a) that could possibly be an effect of an underlying causal structure in accordance with (HC). Analogous considerations apply to list (b). In R1 and R4 of that list A is the only varying factor, R2 and R4 exclude B from being interpretable as an effect, and R3 and R4 refuse C admittance into the set of possible effects due a violation of (HC).

That means, there cannot be a causal structure underlying either list (a) or (b) that would be compatible with (HC). In neither list there is a factor that could be seen as an effect in accordance with (HC), i.e. $\mathbf{W} = \emptyset$. Whenever for *every* factor Z_i contained in the factor frame of a coincidence list \mathcal{C} there are two rows Rk and Rl in \mathcal{C} such that Z_i is the only factor varying in Rk and Rl , the background against which the data in \mathcal{C} is collected cannot be homogeneous, for there is no causal structure that could possibly generate \mathcal{C} and accord with (HC). We shall in this context speak of *inhomogeneous coincidence lists*.

Inhomogeneous coincidence list: A coincidence list \mathcal{C} over a factor frame $\{Z_1, Z_2, \dots, Z_n\}$ is inhomogeneous if for every $Z_i \in \{Z_1, Z_2, \dots, Z_n\}$: \mathcal{C} contains two rows Rk and Rl such that Z_i is the only factor varying in Rk and Rl .

(HC) excludes all inhomogeneous coincidence lists from being processed by CA. It must be emphasized, however, that the homogeneity of coincidence lists is an assumption to which every inference of CA must be relativized. It might well be that a coincidence list which is not inhomogeneous in the sense defined above, as e.g.

list (b) in table 5.6, in fact is the result of an uncontrolled variation of background confounders. In this sense, only a sufficient and no necessary condition for the inhomogeneity of a coincidence list is given above. Causal inferences drawn by CA will always be of the form “Given that (HC) is satisfied, such and such must be the underlying causal structure”. As mentioned in section 5.2, the homogeneity of test situations is never beyond doubt. Nonetheless, assembling the theoretic preconditions of conclusive causal inferences serves the goal of establishing experimental standards, implementable as a gauge for concrete experimental settings and causal inferences.

5.5.4 Identification of Potential Effects

After having clarified the presuppositions on which CA rests, we can now proceed to introduce the inference rules of CA. As anticipated in the previous section, a first algorithmic step consists in parsing through the factor frame of a coincidence list in order to determine which of the factors could possibly operate as effects within the causal structure to be revealed. This step yields a set W of factors whose dependencies on the other factors in the corresponding frame are then successively determined by CA. The identification of potential effects shall not be considered a proper part of CA, for any sort of context dependent empirical information or even prior causal knowledge is allowed to enter the determination of W . For instance, if a factor Z_i is generally instantiated *temporally before* every other factor in an analyzed frame $\{Z_1, \dots, Z_n\}$, Z_i cannot function as an effect within the underlying causal structure. Or *prior causal knowledge* could be available that establishes the members of a proper subset of $\{Z_1, \dots, Z_n\}$ as root factors. In both cases there is no need to integrate respective factors in W . CA does not have to evaluate dependencies among factors that can be excluded from the set of potential effects to begin with. These pragmatic circumstances are not systematizable or at least a systematization shall not be attempted here. Accordingly, no recursively applicable or computable rule can be provided, which essentially is why the determination of W is not seen as a proper part of CA.

Still, the determination of W is not only regulated by spatiotemporal peculiarities of an analyzed process or by prior causal knowledge. As the previous section has shown, factors can be excluded from the set of potential effects based on homogeneity considerations. Test situations recorded in coincidence lists are assumed to be homogeneous in terms of (HC). Now, if a factor Z_i is the only factor varying in two test situations and if, moreover, Z_i were an effect of the underlying causal structure, (HC) would be violated with respect to Z_i . In order for a factor Z_i to be a potential effect, it must not be the case that the corresponding coincidence list contains two rows such that Z_i is the only varying factor in those rows. As a consequence of assuming that the background of coincidence lists is homogeneous, a factor Z_i can only be regarded as a potential effect if Z_i is not the only factor varying in any two test situations. This induces a first necessary condition with respect to membership in W .

A	B	C	A	B	C	A	B	C	A	B	C
1	1	1	0	1	1	1	1	1	1	1	1
1	0	1	0	0	1	1	1	1	1	1	1
1	1	0	0	1	0	1	1	1	0	0	0
1	0	0	0	0	0	1	1	1	0	0	0
(a)			(b)			(c)			(d)		

Tab. 5.8: Four exemplary coincidence lists whose set W of potential effects does not comprise all factors in the respective factor frame. While A is not contained in W relative to lists (a) and (b), $W = \emptyset$ for (c) and (e).

Compatibility with (HC): A factor Z_i contained in a coincidence list C over a factor frame $\{Z_1, \dots, Z_n\}$ is a member of the set W of factors that possibly operate as effects within the causal structure regulating the behavior of $\{Z_1, \dots, Z_n\}$ only if C does not contain two rows Rk and Rl such that Z_i is the only factor varying in the test situations recorded by Rk and Rl .

Furthermore, factors can be excluded from the set of potential effects if they violate causal principles. Consider the lists (a) and (b) in table 5.8. In (a) factor A is universally present, while in (b) A is universally absent. The Principle of Relevance (PR) demands that every causally relevant factor is indispensable for the instantiation of the corresponding effect in at least one test situation, i.e. brings about that effect while all its alternative causes are absent. Such indispensability is only empirically warrantable if there are at least two test situations – one *with* instantiated effect, one *without* – such that the factor under consideration accounts for the difference in regard to instantiations of the effect. The data on the behavior of A recorded both in (a) and in (b), thus, does not conform to (PR), because there are no two test situations as demanded by (PR) in either (a) or (b). It is therefore ascertained on a priori grounds that, if CA assigns a minimal theory to either (a) or (b), A is not contained in effect position in that theory. Accordingly, A does not have to be integrated in W . The following minimality constraint on the variability of members of W amounts to a second necessary condition for membership in W :

Minimal variability of potential effects: A factor Z_i contained in a coincidence list C over a factor frame $\{Z_1, \dots, Z_n\}$ is a member of the set W of factors that possibly operate as effects within the causal structure regulating the behavior of $\{Z_1, \dots, Z_n\}$ only if C records at least one coincidence featuring Z_i and one coincidence featuring \bar{Z}_i .

If this requirement as to the minimal variability of potential effects is applied to the lists (c) and (d) in table 5.8, it is easily seen that – as in case of inhomogeneous coincidence lists – the set of potential effects is empty, $W = \emptyset$. Whenever that is the case, CA does not have to be launched in the first place. An empty set of potential effects is implementable as a sort of termination criterion for CA.

Basically, since **CA** shall be designed to infer causes of both positive and negative factors, **W** may prima facie contain both positive and negative factors. However, section 3.6.4 has shown that to every minimal theory of a positive factor Z_i , there exists an equivalent minimal theory of \overline{Z}_i , and vice versa. Hence, for simplicity's sake, **CA** can be confined to identify minimal theories of either positive factors or their negative counterparts – conducting both at the same time being theoretically possible, but practically unnecessary. For this reason, we stipulate that positive factors only shall be included in **W**.

These considerations taken together yield the following standard as regards the determination of **W**. In order to indicate that the non-computable identification of the set of potential effects is a precondition of launching **CA**, yet not a proper part thereof, it shall be referred to as “step 0*”.

Step 0 – Identification of potential effects:* Given a coincidence list \mathcal{C} over a factor frame $\{Z_1, \dots, Z_n\}$, identify the subset $\mathbf{W} \subseteq \{Z_1, \dots, Z_n\}$ such that for every Z_i : $Z_i \in \mathbf{W}$ iff

- (1) The totality of available information as to the spatiotemporal ordering of the instances of the factors in $\{Z_1, \dots, Z_n\}$ and the available prior causal knowledge about the behavior of the factors in $\{Z_1, \dots, Z_n\}$ does not preclude Z_i to be an effect of the underlying causal structure.
- (2) \mathcal{C} does not contain two rows Rk and Rl such that Z_i is the only factor varying in the test situations recorded by Rk and Rl .
- (3) \mathcal{C} records at least one coincidence featuring Z_i and one coincidence featuring \overline{Z}_i .
- (4) Z_i is a positive factor.

5.5.5 Identification and Minimalization of Sufficient Conditions

After having identified a non-empty set of potential effects, **CA** proper sets in. In a first stage, sufficient conditions for each member of **W** are identified and minimalized. In order to illustrate this first stage, let us look at a concrete example. Assume the coincidence list depicted in table 5.9 to be our input data. None of the factors in our exemplary frame $\{A, B, C, D, E\}$ shall be excluded from effect position by prior causal knowledge or additional information as to the spatiotemporal ordering of the instances of these factors. Nonetheless, even though none of the factors is constantly present or absent, the set of potential effects does not correspond to the factor frame of table 5.9, i.e. $\mathbf{W} \neq \{A, B, C, D, E\}$. For reasons of compatibility with (HC), factors A , B , and D cannot be effects. For each of these factors there is a pair of rows in table 5.9 – $\langle R1, R5 \rangle$ for A , $\langle R1, R3 \rangle$ for B , $\langle R1, R2 \rangle$ for D – such that the respective factor is the only varying factor. Thus, interpreting one of these factors to be an effect of the underlying causal structure would contradict **CA**'s homogeneity assumption. C and E , thus, are the only potential effects of the causal structure generating table 5.9, i.e. $\mathbf{W} = \{C, E\}$. For each of the factors in **W** minimally sufficient conditions are now identified. This is done in four steps:

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
1	1	1	1	1
1	1	1	0	1
1	0	1	1	1
1	0	1	0	1
0	1	1	1	1
0	1	1	0	1
0	0	0	1	1
0	0	0	0	0

Tab. 5.9: Exemplary coincidence list to be analyzed by CA.

(1) A factor $Z_i \in \mathbf{W}$ is selected, (2) sufficient conditions of Z_i are identified, (3) these sufficient conditions are minimalized, (4) the procedure is restarted at (1) by selecting another $Z_j \in \mathbf{W}$ until all factors in \mathbf{W} have been selected. Let us take a detailed look at these four steps.

Step 1 – Selection of a potential effect: Randomly select one factor $Z_i \in \mathbf{W}$ such that Z_i has not been selected in a previous run of steps 1 to 4. Z_i is termed *effect**, the factors in $\{Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n\}$ are referred to as *remainders*.³⁶

Step 2 – Identification of sufficient conditions: Identify all sufficient conditions of the effect* Z_i according to the following rule:

(SUF) A coincidence X_k of remainders is *sufficient* for Z_i iff the input list \mathcal{C} contains at least one row featuring $X_k Z_i$ and no row featuring $X_k \bar{Z}_i$.

The order of selecting effects* in step 1 does not matter, as long as it is guaranteed that eventually all members of \mathbf{W} are selected. The rule (SUF), that assigns sufficient conditions to effects*, contributes to doing justice to the Principle of Relevance (PR) by only recognizing sufficient conditions that are instantiated at least once. Moreover, a coincidence of remainders contained in the input list is not sufficient for a selected effect* if it is also instantiated in combination with the absence of that effect*.

Let us perform these two steps on our example of table 5.9 by first selecting C as effect*. Step 2 identifies three sufficient conditions of C , i.e. there are six coincidences of remainders that conform to (SUF): $ABDE$, $AB\bar{D}\bar{E}$, $\bar{A}\bar{B}DE$, $\bar{A}\bar{B}\bar{D}\bar{E}$, $\bar{A}BDE$, $\bar{A}B\bar{D}\bar{E}$. The first row (R1) of table 5.9 features the coincidence $ABDE$ in combination with C and there is no row such that $ABDE$ is contained therein in combination with \bar{C} . $ABDE$, thus, is a sufficient condition of C according

³⁶ Selected factors are labelled *effects** to indicate that they *possibly* are the effects of the causal structure generating the input list. Effects* do not necessarily turn out to be (actual) effects of the underlying causal structure at the end of a CA-analysis.

to (SUF). Analogous considerations apply to the other sufficient conditions mentioned above: R2 is constituted by $AB\bar{D}E$, R3 by $A\bar{B}DE$, R4 by $A\bar{B}\bar{D}E$, R5 by $\bar{A}BDE$, and R6 features $\bar{A}\bar{B}\bar{D}E$ without either of these conditions being contained in combination with \bar{C} in table 5.9. Thus, each coincidence of remainders listed in the six rows featuring an instance of C constitutes a sufficient condition of C .

Before sufficient conditions of the remaining effect* E are identified, we proceed to minimalize the sufficient conditions of C in step 3.

Step 3 – Minimalization of sufficient conditions: The sufficient conditions of Z_i identified in step 2 are minimalized according to the following rule:

(MSUF) A sufficient condition $Z_1Z_2\dots Z_h$ of Z_i is *minimally* sufficient iff neither $\bar{Z}_1Z_2\dots Z_h$ nor $Z_1\bar{Z}_2\dots Z_h$ nor \dots nor $Z_1Z_2\dots\bar{Z}_h$ are sufficient for Z_i according to (SUF).

Or operationally put:

(MSUF') Given a sufficient condition $Z_1Z_2\dots Z_h$ of Z_i , for every $Z_g \in \{Z_1, Z_2, \dots, Z_h\}$, $h \geq g \geq 1$, and every h -tuple $\langle Z_{1'}, Z_{2'}, \dots, Z_{h'} \rangle$ which is a permutation of the h -tuple $\langle Z_1, Z_2, \dots, Z_h \rangle$: Eliminate Z_g from $Z_1Z_2\dots Z_h$ and check whether $Z_1\dots Z_{g-1}Z_{g+1}\dots Z_h\bar{Z}_i$ is contained in a row of C . If that is the case, re-add Z_g to $Z_1\dots Z_{g-1}Z_{g+1}\dots Z_h$ and eliminate Z_{g+1} ; if that is not the case, proceed to eliminate Z_{g+1} without re-adding Z_g . The result of performing this redundancy check on every factor contained in $Z_1Z_2\dots Z_h$ is a set of minimally sufficient conditions of Z_i .

(MSUF) is nothing but an adaptation of the notion of a minimally sufficient condition as defined in section 3.3 to the context of coincidence lists as processed by CA. (MSUF'), on the other hand, can be seen as an operational expression of the analysis of the notion of a minimally sufficient condition implemented in (MSUF). That means, (MSUF) might be rephrased as follows: A sufficient condition $Z_1Z_2\dots Z_h$ of Z_i is *minimally* sufficient iff it results from an application of (MSUF'). At the price of high computational complexity, the formulation of (MSUF') is kept as simple as possible above. The order in which factors are eliminated from sufficient conditions matters as to the minimalization of such conditions – thus the systematic permutation of elimination orders.³⁷ In many cases, however, it is not necessary to completely permute elimination orders. For instance, assume an h -tuple $T_1 = \langle Z_1, \dots, Z_d, Z_{d+1}, \dots, Z_h \rangle$ has been minimalized by means of (MSUF') up to element Z_d , that minimalization of T_1 can be taken

³⁷ This is an important deviation from the minimalization of sufficient conditions as performed by Ragin's *QCA* algorithm. In the vein of the Quine-McCluskey optimization of truth tables, *QCA* only eliminates conjuncts of a sufficient condition if the latter reduced by the respective conjunct is actually contained in the coincidence list. As will be shown below, this restriction is a serious limitation of the minimizability of sufficient conditions involved in complex causal structures.

over for all h -tuples $T_2 = \langle Z_1, \dots, Z_d, Z_{d+1'}, \dots, Z_{h'} \rangle$ that coincide with T_1 up to element Z_d without reapplying (MSUF') to T_2 . Or suppose it has been found that $X_1 = Z_1 \dots Z_d$ is a minimally sufficient condition of an investigated effect and a sufficient condition $X_2 = Z_1 Z_2 \dots Z_h$ containing $Z_1 \dots Z_d$ is to be minimalized by means of (MSUF'). In that case, it is not necessary to minimalize X_2 by first eliminating any factor contained in X_1 followed by the elimination of any other factor of X_1 , for this elimination order would just yield X_1 again.

Further optimizations of (MSUF') are conceivable, yet are not going to be discussed in the present context. More importantly, the intuition behind (MSUF') can be more colloquially captured: Every factor contained in a sufficient condition of Z_i is to be tested for redundancy by eliminating it from that condition and checking whether the remaining condition still is sufficient for Z_i or not. A sufficient condition of Z_i is minimally sufficient iff every elimination of a factor from that condition results in the insufficiency of the remaining condition.

Performing step 3 on our exemplary case is straightforward. Step 2 yielded six sufficient conditions of C . For simplicity's sake, I only illustrate the minimalization of these six conditions by means of two examples. First, take $ABDE$. That this sufficient condition is not minimally sufficient for C is seen by removing, say, D and finding that ABE itself is sufficient for C , for table 5.9 does not contain a row featuring $ABE\bar{C}$. ABE still is not minimally sufficient. For instance, both B and E can be removed without sufficiency being lost. There is no row in 5.9 featuring $A\bar{C}$, which induces that A is sufficient and, since it is a single factor, *minimally* sufficient for C .³⁸ There are other ways to further minimalize ABE : A removal of A and E still yields a sufficient condition of C . There is no row in 5.9 featuring $B\bar{C}$. Therefore B is minimally sufficient for C . Second, let us look at the second sufficient condition of C identified by (SUF). $AB\bar{D}E$ is not minimally sufficient because AB can be removed without sufficiency for C being lost. There is no row in 5.9 featuring $\bar{C}\bar{D}E$, which induces that $\bar{D}E$ is sufficient for C . If $\bar{D}E$ is further reduced, sufficiency is lost. R7 features $\bar{C}E$ and R8 $\bar{C}\bar{D}$, which amounts to neither E nor \bar{D} being sufficient for C . $\bar{D}E$, hence, is minimally sufficient for C . Minimalizing the other sufficient conditions of C by analogously implementing (MSUF') does not yield any further minimally sufficient conditions. All in all, therefore, minimalizing the sufficient conditions of C generates the following three minimally sufficient conditions: A , B , and $\bar{D}E$.

After having identified the minimally sufficient conditions of a first factor $Z_i \in W$, the same needs to be done for all other effects*. We thus need a loop that brings CA back to step 1, if not all factors in W have been assigned minimally sufficient conditions yet.

Step 4 – (MSUF)-Loop: If all $Z_i \in W$ have been selected as effects* proceed to step 5, otherwise go back to step 1.

³⁸ Single factors that are sufficient conditions are minimally sufficient conditions – as shown in section 3.3, p. 96.

Applying this loop to our example yields seven sufficient conditions for E . Each row featuring E comprises a sufficient condition of remainders: $ABCD$, $ABC\bar{D}$, $A\bar{B}CD$, $A\bar{B}C\bar{D}$, $\bar{A}BCD$, $\bar{A}BC\bar{D}$, $\bar{A}\bar{B}CD$. For example, R2 of table 5.9 is constituted by $ABC\bar{D}$ and there is no row featuring $ABC\bar{D}$ along with \bar{E} , or R3 comprises $A\bar{B}CD$ and no row in 5.9 contains $A\bar{B}CD$ in combination with \bar{E} . The sufficiency of the other conditions is analogously demonstrated. Employing (MSUF) or (MSUF') to minimize these conditions brings forth four minimally sufficient conditions of E : A , B , C , and D . Table 5.9 contains no rows featuring either $A\bar{E}$, $B\bar{E}$, $C\bar{E}$, or $D\bar{E}$.

As an overall result of performing the first stage (steps 1 to 4) of CA on our exemplary case, we have thus identified the following minimally sufficient conditions of the corresponding factors in W .

Result after 4 steps:

$Z_i \in W$	minimally sufficient conditions
C	$A, B, \bar{D}E$
E	A, B, C, D

In light of having completed the identification of minimally sufficient conditions for all factors in W , the (MSUF)-loop now urges us to proceed to step 5.³⁹

5.5.6 Identification and Minimalization of Necessary Conditions

As the Manchester-Hooters counterexample against Mackie's (1974) INUS-theory of causation⁴⁰ demonstrates, minimally sufficient conditions are not generally causally interpretable. Only minimally sufficient conditions that are moreover non-redundant parts of minimally necessary conditions are amenable to a causal

³⁹ Emphasis must be put on a major difference between Ragin's (1987) *QCA* algorithm and CA that becomes apparent at this point. By presupposing that potential causes of an investigated causal structure are independent (IND) and are, thus, co-instantiable in all logically possible combinations *QCA* can draw on the well-known Quine-McCluskey optimization of truth tables in order to minimize sufficient conditions (cf. Quine (1952), Quine (1959)). As soon as (IND) is dropped, however, Quine-McCluskey optimization no longer eliminates all redundancies. Our exemplary coincidence list in table 5.9 features a dependency among $A \vee B$ and C . There is no row in 5.9 reporting a coincidence of, say, A and \bar{C} . Quine-McCluskey optimization, however, only eliminates redundant conjuncts of sufficient conditions if a respective truth table contains two rows which differ only with respect to presence and absence of that conjunct. Thus, minimalizing the sufficient conditions of E in table 5.9 along the lines of Quine-McCluskey would not identify, say, A as minimally sufficient condition of E , notwithstanding the fact that table 5.9 does not contain a coincidence of A and \bar{E} . Rendering coincidence lists generated by complex causal structures amenable to a Boolean analysis, accordingly, calls for a custom-built minimalization procedure that differs from a standard Quine-McCluskey optimization insofar as it systematically tests conjuncts Z_i of a sufficient condition X_i for eliminability, irrespective of whether the corresponding coincidence list contains another sufficient condition X_j that only differs from X_i with respect to presence and absence of Z_i .

⁴⁰ Cf. Mackie (1974), Baumgartner and Graßhoff (2004), ch. 5, see also section 3.4 above.

interpretation. After having identified minimally sufficient conditions, we thus now proceed to first form necessary conditions of the effects* from their minimally sufficient conditions and then minimalize these necessary conditions. Since factor frames processed by CA are incomplete with respect to underlying causal structures, i.e. there supposedly will always be many causally relevant factors not listed in input lists, effects* can only be assigned necessary conditions relative to the homogeneous backgrounds of corresponding coincidence lists. This is easily accomplished by disjunctively combining the minimally sufficient conditions of each effect*, yielding one necessary condition relative to an input list C and its background for each factor $Z_i \in \mathbf{W}$.

Step 5 – Identification of necessary conditions: Identify a necessary condition of each effect* Z_i by disjunctively concatenating Z_i 's minimally sufficient conditions according to the following rule:

(NEC) A disjunction $X_1 \vee X_2 \vee \dots \vee X_h$ of minimally sufficient conditions of Z_i is *necessary* for Z_i iff C contains no row featuring Z_i in combination with $\neg(X_1 \vee X_2 \vee \dots \vee X_h)$, i.e. no row comprising $\overline{X_1 X_2 \dots X_h} Z_i$.

Performed on our example, step 5 issues the following necessary conditions for C and E :⁴¹

$$C \mapsto A \vee B \vee \overline{D}E \quad (5.6)$$

$$E \mapsto A \vee B \vee C \vee D \quad (5.7)$$

Within the homogeneous background of table 5.9, $A \vee B \vee \overline{D}E$ is necessary for C , and $A \vee B \vee C \vee D$ for E . That means, there is no row in 5.9 featuring C in combination with neither A nor B nor $\overline{D}E$. Whenever C is instantiated, there is also an instance of at least one of its minimally sufficient conditions. Similarly for E : No row of 5.9 records a coincidence of E with neither an instance of A nor B nor C nor D . E is always instantiated in combination with one of its minimally sufficient conditions.

Such as to determine whether the minimally sufficient conditions assigned to the effects* at the end of the previous section in fact are non-redundant parts of necessary conditions, these necessary conditions have to be minimalized.

Step 6 – Minimalization of necessary conditions: The necessary conditions of every $Z_i \in \mathbf{W}$ identified in step 5 are minimalized according to the following rule:

(MNEC) A necessary condition $X_1 \vee X_2 \vee \dots \vee X_h$ of Z_i is *minimally* necessary iff neither $X_2 \vee X_3 \vee \dots \vee X_h$ nor $X_1 \vee X_3 \vee \dots \vee X_h$ nor \dots nor $X_1 \vee X_2 \vee \dots \vee X_{h-1}$ is necessary for Z_i according to (NEC).

Or operationally put:

⁴¹ For a definition of “ \mapsto ” cf. section 3.2, p. 94.

(MNEC') Given a necessary condition $X_1 \vee X_2 \vee \dots \vee X_h$ of Z_i , for every $X_g \in \{X_1, X_2, \dots, X_h\}$, $h \geq g \geq 1$, and every h -tuple $\langle X_{1'}, X_{2'}, \dots, X_{h'} \rangle$ which is a permutation of the h -tuple $\langle X_1, X_2, \dots, X_h \rangle$: Eliminate X_g from $X_1 \vee X_2 \vee \dots \vee X_h$ and check whether there is a row in \mathcal{C} featuring Z_i in combination with $\neg(X_1 \vee \dots \vee X_{g-1} \vee X_{g+1} \vee \dots \vee X_h)$, i.e. a row comprising $\overline{X_1 \dots X_{g-1} X_{g+1} \dots X_h} Z_i$. If that is the case, re-add X_g to $X_1 \vee \dots \vee X_{g-1} \vee X_{g+1} \vee \dots \vee X_h$ and eliminate X_{g+1} ; if that is not the case, proceed to eliminate X_{g+1} without re-adding X_g . The result of performing this redundancy check on every minimally sufficient condition contained in $X_1 \vee X_2 \vee \dots \vee X_h$ is a set of minimally necessary conditions of Z_i .

In analogy to (MSUF), (MNEC) is nothing but an adaptation of the notion of a minimally necessary condition as defined in section 3.4 to the context of coincidence lists. (MNEC'), on the other hand, can be seen as an operational expression of the analysis of the notion of a minimally necessary condition implemented in (MNEC). That means, (MNEC) might be rephrased as follows: A necessary condition $X_1 \vee X_2 \vee \dots \vee X_h$ is *minimally* necessary iff it results from an application of (MNEC'). The formulation of (MNEC') has been kept as simple as possible at the expense of its computational complexity. Analogous optimizations as in case of (MSUF')⁴² are possible with respect to (MNEC'). The intuition behind (MNEC') can also be more colloquially captured: Every minimally sufficient condition contained in a necessary condition of Z_i is to be tested for redundancy by eliminating it from that condition and checking whether the remaining condition still is necessary for Z_i or not. A necessary condition of Z_i is minimally necessary iff every elimination of a minimally sufficient condition from that necessary condition results in the loss of necessity of the remaining condition.

Let us illustrate the minimalization of necessary conditions by first performing step 6 on the necessary condition of C specified in (5.6). $A \vee B \vee \overline{DE}$ is not minimally necessary for C , because it contains a necessary proper part: $A \vee B$. Whenever C is instantiated in table 5.9, there is an instance of either A or B . 5.9 does not contain a row featuring \overline{ABC} . \overline{DE} does not amount to a non-redundant part of a minimally necessary condition, for whenever \overline{DE} is instantiated in combination with C , there also is an instance of $A \vee B$. The same results from applying (MNEC') to $A \vee B \vee \overline{DE}$. When eliminating A we find that the rest is no longer necessary for C , because R3 of table 5.9 features \overline{BDEC} , or more specifically \overline{BDEC} . Hence, A is re-added. The same is found upon removing B . R5 features \overline{ADEC} or \overline{ADEC} , respectively. Removing \overline{DE} , however, does not result in a loss of necessity. Therefore, \overline{DE} is not re-added. The necessary condition of E exhibited in (5.7) neither amounts to a minimally necessary condition. $A \vee B \vee C \vee D$ not only contains one but two necessary proper parts: $C \vee D$ and $A \vee B \vee D$. There is no row in 5.9 featuring \overline{CDE} or \overline{ABDE} . Whenever E is instantiated, there is

⁴² Cf. p. 238 above.

an instance of $C \vee D$ and of $A \vee B \vee D$. These two ways to minimize (5.7) stem from the fact that there are biconditional dependencies among the minimally sufficient conditions of E . Within the homogeneous background of table 5.9, C is instantiated if and only if $A \vee B$ is instantiated.

Result after 6 steps:

$Z_i \in \mathbf{W}$	minimally necessary conditions
C	$A \vee B$
E	$A \vee B \vee D, C \vee D$

5.5.7 Framing Minimal Theories

Step 6 of CA yields a set of minimally necessary disjunctions of minimally sufficient conditions for each $Z_i \in \mathbf{W}$. We have thus come close to assigning a minimal theory to table 5.9. A simple minimal theory of a factor Z_i in fact is nothing but a double-conditional whose antecedent is constituted by a minimally necessary disjunction of minimally sufficient conditions of Z_i and whose consequent comprises Z_i itself. The result of step 6 allows for framing one simple minimal theory for C and two for E . Relative to the homogeneous background of table 5.9, these minimal theories can be straightforwardly expressed by means of propositional logic:

$$A \vee B \leftrightarrow C \quad (5.8)$$

$$A \vee B \vee D \leftrightarrow E \quad (5.9)$$

$$C \vee D \leftrightarrow E \quad (5.10)$$

However, apart from the specific context of table 5.9, it is certainly not the case that A and B are themselves sufficient for C or C and D for E . Moreover, there may well be further minimally sufficient conditions for both C and E . Therefore, suspending the relativization to the background of 5.9 and expressing these dependencies in their general double-conditional and context independent form leads to:

$$AX_1 \vee BX_2 \vee Y_C \Rightarrow C \quad (5.11)$$

$$AX_1 \vee BX_2 \vee DX_3 \vee Y_E \Rightarrow E \quad (5.12)$$

$$CX_1 \vee DX_2 \vee Y_E \Rightarrow E \quad (5.13)$$

C and E have a non-empty intersection of minimally sufficient conditions. Correspondingly, the simple minimal theories of C and E share a number of common factors. The causal structure regulating the behavior of E is not independent of the structure behind the behavior of C . The behavior of the factors in table 5.9, thus, is regulated by a complex causal structure. In order to determine what that

structure looks like, the simple minimal theories of C and E are to be conjunctively combined to form a complex theory. Here an ambiguity emerges that we have already encountered in chapter 4. (5.12) and (5.13) – if causally interpreted – identify different direct causal relevancies for E . While according to (5.12) A and B are directly causally relevant to E , (5.13) instead holds C to be directly relevant to E . The coincidences in table 5.9 are either generated by a causal chain, such that A and B are parts of alternative causes of C while C and D are contained in alternative causes of E , or they are generated by an epiphenomenon, such that A and B are parts of alternative causes of C while A , B , and D are contained in alternative causes of E . The ambiguous minimalization of E 's necessary condition is nothing but the inferential variant of what has been termed the *chain-problem* in the previous chapter. Prima facie, the list in table 5.9 is underdetermined as to whether its coincidences are the result of a chain or an epiphenomenon. However, the coincidences in 5.9 not only induce an ambiguity with respect to the minimalization of E 's necessary condition, they moreover reveal that C and E are *entangled* factors.⁴³ Every part of a minimally sufficient condition of C is also part of a minimally sufficient condition of E . In chapter 4 we have seen that entanglements that resist all factor frame extensions are to be causally interpreted. Of course, the factor frame involved in table 5.9 cannot be seen as fully expanded as regards underlying causal structures. Nonetheless, in line with the solution of the chain-problem developed in chapter 4, we shall at this point – pending a complete expansion of the corresponding factor frame – give preference to a causal interpretation of entanglements. As long as entanglements have not been suspended by the introduction of additional factors, CA shall causally interpret entanglements. That means that complex minimal theories are to be built up from simple theories such that for every i , $1 \leq i < n$, in a sequence of entangled factors Z_1, \dots, Z_n , $n \geq 2$: Z_i is contained in the antecedent of the simple minimal theory of Z_{i+1} . This requirement is mirrored by the notion of a complex minimal theory developed in section 4.5.2. Hence, the notion of a complex minimal theory itself makes sure that not any conjunction of the biconditionals (5.8), (5.9), and (5.10) or their first-order correlates amounts to a complex minimal theory. Of (5.14) and (5.15) only (5.14) constitutes a complex minimal theory. (5.15) cannot be seen as complex theory for it does not conform to the syntactical constraints imposed on entangled factors in chapter 4.

$$\begin{aligned} & (A \vee B \leftrightarrow C) \wedge (C \vee D \leftrightarrow E) \\ & (AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (CX_3 \vee DX_4 \vee Y_E \Rightarrow E) \end{aligned} \quad (5.14)$$

$$\begin{aligned} & (A \vee B \leftrightarrow C) \wedge (A \vee B \vee D \leftrightarrow E) \\ & (AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (AX_1 X_3 \vee BX_2 X_3 \vee Y_C X_3 \vee DX_4 \vee Y_E \Rightarrow E) \end{aligned} \quad (5.15)$$

All in all, in the remaining step of CA, minimal theories are framed from the minimally necessary disjunctions of minimally sufficient conditions identified for

⁴³ Cf. chapter 4, especially section 4.3.

each $Z_i \in \mathbf{W}$ in step 6. This is done by means of a twofold procedure: First, simple minimal theories are formed for each $Z_i \in \mathbf{W}$, and second, if the minimal theories Φ and Ψ of two different factors in \mathbf{W} have a non-empty intersection of factors, Φ and Ψ are combined to form the complex minimal theory $\Phi \wedge \Psi$, such that $\Phi \wedge \Psi$ conforms to the requirements imposed on the notion of a complex minimal theory at the end of the previous chapter.⁴⁴

Step 7 – Framing minimal theories: The minimally necessary disjunctions of minimally sufficient conditions of each $Z_i \in \mathbf{W}$ identified in step 6 are assembled to minimal theories as follows:

- (1) For each $Z_i \in \mathbf{W}$ and each minimally necessary disjunction $X_1 \vee X_2 \vee \dots \vee X_h$, $h \geq 2$,⁴⁵ of minimally sufficient conditions of Z_i : form a simple minimal theory Ψ of Z_i by making $X_1 \vee X_2 \vee \dots \vee X_h$ the antecedent of a double-conditional and Z_i its consequent: $X_1 \vee X_2 \vee \dots \vee X_h \Rightarrow Z_i$.
- (2) Conjunctively combine two simple minimal theories Φ and Ψ to the complex minimal theory $\Phi \wedge \Psi$ iff Φ and Ψ conform to the following conditions:
 - (a) at least one factor in Φ is part of Ψ ;
 - (b) Φ and Ψ do not have an identical consequent;
 - (c) for every i , $1 \leq i < n$, in a sequence of entangled factors Z_1, \dots, Z_n , $n \geq 2$: Z_i is contained in the antecedent of the simple minimal theory of Z_{i+1} .

Applied to our example, step 7 assigns the following complex minimal theory to table 5.9:

$$(AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (CX_3 \vee DX_4 \vee Y_E \Rightarrow E) \quad (5.16)$$

Result after 7 steps:

$Z_i \in \mathbf{W}$	minimal theory
C	$(AX_1 \vee BX_2 \vee Y_C \Rightarrow C) \wedge (CX_3 \vee DX_4 \vee Y_E \Rightarrow E)$
E	

All that remains to be done now on the way to determine the causal structure underlying the list in table 5.9 is to causally interpret (5.16), which, of course, is straightforward.

⁴⁴ Cf. section 4.5.2, p. 171.

⁴⁵ The constraint as to a minimum of two alternative minimally sufficient conditions for each effect* does justice to the minimal complexity of a causal structure required such that its direction is identifiable (cf. section 3.6.2 above). See also section 5.7 below.

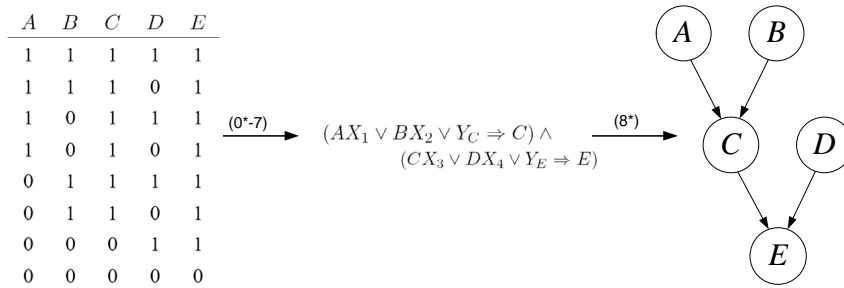


Fig. 5.7: CA identifies the causal structure underlying our exemplary input list by assigning a minimal theory to that list. Steps 0* to 7 are concerned with the specification of a minimal theory, step 8* causally interprets that theory.

5.5.8 Causal Interpretation

After having assigned a minimal theory to a coincidence list, the by far most intricate hurdles on the way to uncovering the causal structure behind that list have been overcome. As we have seen on several occasions, there exists a plain syntactical convention as regards the causal interpretation of minimal theories. Minimal theories render causal structures syntactically transparent:

Step 8 – Causal interpretation:* Disjuncts in the antecedent of simple minimal theories are to be interpreted as alternative (complex) causes of the factor in the consequent. Conjuncts constituting such disjuncts correspond to non-redundant parts of complex causes. Triples of factors $\langle Z_h, Z_i, Z_j \rangle$, such that Z_h appears in the antecedent of a minimal theory of Z_i and Z_i is part of a minimal theory of Z_j , are to be interpreted as causal chains.

This interpretation rule is not to be seen as part of CA proper. Nonetheless, it fulfills an essential function on the way to a causal inference. For this reason, the rule concerning causal interpretation is starred.

CA thus determines the coincidences in our exemplary table 5.9 to be the result of the causal chain depicted in figure 5.7. *A* and *B* are parts of alternative causes of *C* while *C* and *D* are contained in alternative causes of *E*. Thereby, *A* and *B* are moreover rendered indirectly causally relevant to *E*. Steps 0* to 7 assign a minimal theory to an input list and step 8* causally interprets that theory.

5.5.9 Summary of the Algorithm

Step 0 – Identification of potential effects:* Given a coincidence list \mathcal{C} over a factor frame $\{Z_1, \dots, Z_n\}$, identify the subset $\mathbf{W} \subseteq \{Z_1, \dots, Z_n\}$ such that for every Z_i : $Z_i \in \mathbf{W}$ iff

- (1) The totality of available information as to the spatiotemporal ordering of the instances of the factors in $\{Z_1, \dots, Z_n\}$ and the available prior

causal knowledge about the behavior of the factors in $\{Z_1, \dots, Z_n\}$ does not preclude Z_i to be an effect of the underlying causal structure.

- (2) \mathcal{C} does not contain two rows Rk and Rl such that Z_i is the only factor varying in the test situations recorded by Rk and Rl .
- (3) \mathcal{C} records at least one coincidence featuring Z_i and one coincidence featuring $\overline{Z_i}$.
- (4) Z_i is a positive factor.

Step 1 – Selection of a potential effect: Randomly select one factor $Z_i \in \mathbf{W}$ such that Z_i has not been selected in a previous run of steps 1 to 4. Z_i is termed *effect**, the remaining factors in $\{Z_1, \dots, Z_n\}$ are referred to as *remainders*.

Step 2 – Identification of sufficient conditions: Identify all sufficient conditions of the effect* Z_i according to the following rule:

- (SUF) A coincidence X_k of remainders is *sufficient* for Z_i iff \mathcal{C} contains at least one row featuring $X_k Z_i$ and no row featuring $X_k \overline{Z_i}$.

Step 3 – Minimalization of sufficient conditions: The sufficient conditions of Z_i identified in step 2 are minimalized according to the following rule:

- (MSUF) A sufficient condition $Z_1 Z_2 \dots Z_h$ of Z_i is *minimally* sufficient iff neither $\overline{Z_1} Z_2 \dots Z_h$ nor $Z_1 \overline{Z_2} \dots Z_h$ nor ... nor $Z_1 Z_2 \dots \overline{Z_h}$ are sufficient for Z_i according to (SUF).

Step 4 – (MSUF)-Loop: If all $Z_i \in \mathbf{W}$ have been selected as effects* proceed to step 5, otherwise go back to step 1.

Step 5 – Identification of necessary conditions: Identify a necessary condition of each effect* Z_i by disjunctively concatenating Z_i 's minimally sufficient conditions according to the following rule:

- (NEC) A disjunction $X_1 \vee X_2 \vee \dots \vee X_h$ of minimally sufficient conditions of Z_i is *necessary* for Z_i iff \mathcal{C} contains no row featuring Z_i in combination with $\neg(X_1 \vee X_2 \vee \dots \vee X_h)$, i.e. no row comprising $\overline{X_1} \overline{X_2} \dots \overline{X_h} Z_i$.

Step 6 – Minimalization of necessary conditions: The necessary conditions of every $Z_i \in \mathbf{W}$ identified in step 5 are minimalized according to the following rule:

- (MNEC) A necessary condition $X_1 \vee X_2 \vee \dots \vee X_h$ of Z_i is *minimally* necessary iff neither $X_2 \vee X_3 \vee \dots \vee X_h$ nor $X_1 \vee X_3 \vee \dots \vee X_h$ nor ... nor $X_1 \vee X_2 \vee \dots \vee X_{h-1}$ is necessary for Z_i according to (NEC).

Step 7 – Framing minimal theories: The minimally necessary disjunctions of minimally sufficient conditions of each $Z_i \in \mathbf{W}$ identified in step 6 are assembled to minimal theories as follows:

- (1) For each $Z_i \in \mathbf{W}$ and each minimally necessary disjunction $X_1 \vee X_2 \vee \dots \vee X_h$, $h \geq 2$, of minimally sufficient conditions of Z_i : form a simple minimal theory Ψ of Z_i by making $X_1 \vee X_2 \vee \dots \vee X_h$ the antecedent of a double-conditional and Z_i its consequent: $X_1 \vee X_2 \vee \dots \vee X_h \Rightarrow Z_i$.
- (2) Conjunctively combine two simple minimal theories Φ and Ψ of to the complex minimal theory $\Phi \wedge \Psi$ iff Φ and Ψ conform to the following conditions:
 - (a) at least one factor in Φ is part of Ψ ;
 - (b) Φ and Ψ do not have an identical consequent;
 - (c) for every i , $1 \leq i < n$, in a sequence of entangled factors Z_1, \dots, Z_n , $n \geq 2$: Z_i is contained in the antecedent of the simple minimal theory of Z_{i+1} .

*Step 8** – *Causal interpretation*: Disjuncts in the antecedent of simple minimal theories are to be interpreted as alternative (complex) causes of the factor in the consequent. Conjuncts constituting such disjuncts correspond to non-redundant parts of complex causes. Triples of factors $\langle Z_h, Z_i, Z_j \rangle$, such that Z_h appears in the antecedent of a minimal theory of Z_i and Z_i is part of a minimal theory of Z_j , are to be interpreted as causal chains.

5.6 Further Examples

After having completely laid out CA and after having tested its performance with respect to a first concrete coincidence list, CA shall now be applied to two further examples. Consider first the list in table 5.10. It covers the same factor frame as table 5.9 and only differs from the latter in regard to four rows: R2, R4, R6, R7. In order to determine the set \mathbf{W} of potential effects, it again is assumed that no factor in $\{A, B, C, D, E\}$ is excluded from effect position by prior causal knowledge or spatiotemporal constraints. For reasons of compatibility with (HC), however,

A	B	C	D	E
1	1	1	1	1
1	1	0	1	1
1	0	1	1	1
1	0	0	1	0
0	1	1	1	1
0	1	0	1	1
0	0	1	0	1
0	0	0	0	0

Tab. 5.10: A second exemplary coincidence list over the same factor frame as the list in table 5.9.

factors A , B , and C cannot be effects. For each of these factors there is a pair of rows in table 5.10 – $\langle R1, R5 \rangle$ for A , $\langle R1, R3 \rangle$ for B , $\langle R1, R2 \rangle$ for C – such that the respective factor is the only varying factor. D and E , thus, are the potential effects of the causal structure generating table 5.10, i.e. $\mathbf{W} = \{D, E\}$.

Performing steps 2 and 3 on D and E yields the following:

Sufficient conditions of D : $ABCE, AB\bar{C}E, A\bar{B}CE, A\bar{B}\bar{C}E, \bar{A}BCE, \bar{A}B\bar{C}E$.

Minimally sufficient conditions of D : $A, B, \bar{C}E$.

Sufficient conditions of E : $ABCD, AB\bar{C}D, A\bar{B}CD, \bar{A}BCD, \bar{A}B\bar{C}D, \bar{A}\bar{B}\bar{C}D$.

Minimally sufficient conditions of E : $B, C, \bar{A}D$.

After having identified minimally sufficient conditions, CA proceeds to first form and then minimize necessary conditions for each effect*.

Necessary condition of D : $A \vee B \vee \bar{C}E$.

Minimally necessary condition of D : $A \vee B$.

Necessary condition of E : $B \vee C \vee \bar{A}D$.

Minimally necessary condition of E : $B \vee C$.

The minimally sufficient conditions $\bar{C}E$ of D and $\bar{A}D$ of E are not part of minimally necessary conditions of D and E , for whenever they are instantiated, there is an instance of another disjunct in the corresponding necessary conditions. The two conditions thus are redundant within their necessary conditions. Finally, CA frames one simple minimal theory for D and E each:

$$AX_1 \vee BX_2 \vee Y_D \Rightarrow D \quad (5.17)$$

$$BX_3 \vee CX_4 \vee Y_E \Rightarrow E \quad (5.18)$$

(5.17) and (5.18) have one factor in common – B – while none are entangled. Hence, the two simple minimal theories can unambiguously be conjunctively joined to constitute a complex theory representing the causal structure generating table 5.10.

$$(AX_1 \vee BX_2 \vee Y_D \Rightarrow D) \wedge (BX_3 \vee CX_4 \vee Y_E \Rightarrow E) \quad (5.19)$$

Accordingly, this input list is the result of an epiphenomenon as depicted in figure 5.8.

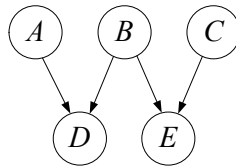


Fig. 5.8: Causal structure underlying the list in table 5.10.

#	A	B	C	D	E	F	G	H	I
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	0
3	1	1	1	1	1	1	1	0	1
4	1	1	1	1	1	1	1	0	0
5	1	1	1	1	1	0	1	1	1
6	1	1	1	1	1	0	1	1	0
7	1	1	1	1	1	0	1	0	1
8	1	1	1	1	1	0	1	0	0
9	1	1	1	0	1	1	1	1	1
10	1	1	1	0	1	1	1	1	0
11	1	1	1	0	1	1	1	0	1
12	1	1	1	0	1	1	1	1	0
13	1	1	1	0	1	0	1	1	1
14	1	1	1	0	1	0	1	1	0
15	1	1	1	0	1	0	1	0	1
16	1	1	1	0	1	0	1	0	0
17	1	0	1	1	1	1	1	1	1
18	1	0	1	1	1	1	1	1	0
19	1	0	1	1	1	1	1	0	1
20	1	0	1	1	1	1	1	0	0
21	1	0	1	1	1	0	1	1	1
22	1	0	1	1	1	0	1	1	0
23	1	0	1	1	1	0	1	0	1
24	1	0	1	1	1	0	1	0	0
25	1	0	0	0	1	1	1	1	1
26	1	0	0	0	1	1	1	1	0
27	1	0	0	0	1	1	1	0	1
28	1	0	0	0	1	1	1	0	0
29	1	0	0	0	0	0	1	1	1
30	1	0	0	0	0	0	0	1	0
31	1	0	0	0	0	0	0	0	1
32	1	0	0	0	0	0	0	0	0
33	0	1	1	1	1	1	1	1	1
34	0	1	1	1	1	1	1	1	0
35	0	1	1	1	1	1	1	0	1
36	0	1	1	1	1	1	1	1	0
37	0	1	1	1	1	0	1	1	1
38	0	1	1	1	1	0	1	1	0
39	0	1	1	1	1	0	1	0	1
40	0	1	1	1	1	0	1	0	0
41	0	1	0	0	1	1	1	1	1
42	0	1	0	0	1	1	1	1	0
43	0	1	0	0	1	1	1	0	1
44	0	1	0	0	1	1	1	0	0
45	0	1	0	0	0	0	1	1	1
46	0	1	0	0	0	0	0	1	0
47	0	1	0	0	0	0	0	0	1
48	0	1	0	0	0	0	0	0	0
49	0	0	1	1	1	1	1	1	1
50	0	0	1	1	1	1	1	1	0
51	0	0	1	1	1	1	1	0	1
52	0	0	1	1	1	1	1	0	0
53	0	0	1	1	1	0	1	1	1
54	0	0	1	1	1	0	1	1	0
55	0	0	1	1	1	0	1	0	1
56	0	0	1	1	1	0	1	0	0
57	0	0	0	0	1	1	1	1	1
58	0	0	0	0	1	1	1	1	0
59	0	0	0	0	1	1	1	0	1
60	0	0	0	0	1	1	1	0	0
61	0	0	0	0	0	0	1	1	1
62	0	0	0	0	0	0	0	1	0
63	0	0	0	0	0	0	0	0	1
64	0	0	0	0	0	0	0	0	0

Tab. 5.11: A third exemplary coincidence list over an extended factor frame.

A considerably more complex example is provided in table 5.11. This coincidence list records the empirically possible coincidences of the nine factors in the frame $\{A, B, C, D, E, F, G, H, I\}$. The list contains a total of 65 rows, which, for purpose of easy reference, are numbered in the first column. In order to determine the set W of potential effects, it is – once again – presumed that no factor in $\{A, B, C, D, E, F, G, H, I\}$ is excluded from effect position by prior causal knowledge or additional information as to the spatiotemporal ordering of the instances of these factors. For reasons of compatibility with (HC), factors $A, B, D, F, H,$ and I cannot be effects. For each of these factors there is a pair of rows in table 5.11 such that the respective factor is the only varying factor relative to that pair. $C, E,$ and $G,$ hence, are the potential effects of the causal structure generating our third exemplary input list, i.e. $W = \{C, E, G\}$.

Step 2 identifies 40 and more sufficient conditions of each factor in W . Listing all of these conditions here would neither be transparent nor serviceable to the purposes at hand. I therefore only mention the minimally sufficient conditions of $C, E,$ and G resulting from performing step 3 of CA on these effects*.

Minimally sufficient conditions of C: $AB, D, E\bar{F}, \bar{F}G\bar{I}, \bar{F}G\bar{H}.$

Minimally sufficient conditions of E: $AB, C, D, F, G\bar{I}, G\bar{H}.$

Minimally sufficient conditions of G: $AB, C, D, E, F, HI.$

For reasons of brevity, I shall only illustrate how CA establishes AB and C to be minimally sufficient for G . Consider, for instance, R7 of table 5.11. R7 features the coincidence $ABCDEF\bar{H}I$ in combination with G . This coincidence amounts to a sufficient condition of G , because no row of 5.11 is constituted by $ABCDEF\bar{G}HI$. Furthermore, a detailed look at the first 16 rows reveals that whenever ABC is instantiated, so is G . Thus, a removal of $DEF\bar{H}I$ from the sufficient condition in R7 does not lead to a loss of sufficiency. ABC still is not minimally sufficient, for there is neither a row featuring $AB\bar{G}$ nor one comprising $C\bar{G}$ in table 5.11. However, there are rows, e.g. R30-R32, featuring $A\bar{G}$ and, e.g. R46-R48, featuring $B\bar{G}$. That means, ABC can be subdivided into two minimally sufficient conditions of G : AB and C . The other minimal sufficiency relationships are analogously established.

The disjunction of their minimally sufficient conditions constitutes a necessary condition of each factor in W as required by step 5.

Necessary condition of C: $AB \vee D \vee E\bar{F} \vee \bar{F}G\bar{I} \vee \bar{F}G\bar{H}.$

Necessary condition of E: $AB \vee C \vee D \vee F \vee G\bar{I} \vee G\bar{H}.$

Necessary condition of G: $AB \vee C \vee D \vee E \vee F \vee HI.$

None of these necessary conditions is minimally necessary. Each contains necessary proper parts. In step 6 CA minimalizes these conditions as follows:

Minimally necessary condition of C: $AB \vee D.$

Minimally necessary conditions of E: $AB \vee D \vee F, C \vee F.$

Minimally necessary conditions of G: $AB \vee D \vee F \vee HI, C \vee F \vee HI, E \vee HI.$

Take, for instance, the necessary condition of C mentioned above. There is no row in table 5.11 that contains an instance of C , but neither one of AB nor one of D . Moreover, $AB \vee D$ is the only minimally necessary proper part of $AB \vee D \vee E\bar{F} \vee \bar{F}G\bar{I} \vee \bar{F}G\bar{H}$. Assume, AB were removed from that necessary condition. In R12 none of the remaining disjuncts is instantiated, but there is an instance of C . Similarly for D : Removing D results in a loss of necessity, for e.g. R17 features C along with none of the remaining disjuncts. Minimalizing the necessary conditions of E and G , however, yields ambiguities. For example, there is no row in 5.11 recording a coincidence of G with neither AB nor D nor F nor HI . The same holds in case of G and $C \vee F \vee HI$ or $E \vee HI$, respectively. As soon as any of these disjuncts is removed, necessity is lost.

Prior to determining which minimal theories CA frames with these minimally necessary conditions, entanglements have to be identified. In fact, there are entangled factors in table 5.11. Every part of a minimally sufficient condition of C is also part of a minimally sufficient condition of E , and every part of a minimally sufficient condition of E is also contained in the minimally sufficient conditions of G . Table 5.11 features a sequence of entangled factors: C, E, G . Therefore, in a complex minimal theory representing the causal structure underlying the behavior of these factors, C appears in the antecedent of the simple minimal theory of E , which, in turn, is part of the antecedent of the simple minimal theory of G . If, furthermore, the relativization to the causal background of the list in table 5.11 is suspended, these considerations yield the following complex minimal theory:

$$(ABX_1 \vee DX_2 \vee Y_C \Rightarrow C) \wedge (CX_3 \vee FX_4 \vee Y_E \Rightarrow E) \wedge (EX_5 \vee HIX_6 \vee Y_G \Rightarrow G). \quad (5.20)$$

CA thus assigns (5.20) to the coincidence list in table 5.11. The causal interpretation of (5.20) is at hand: The coincidences recorded by table 5.11 are generated

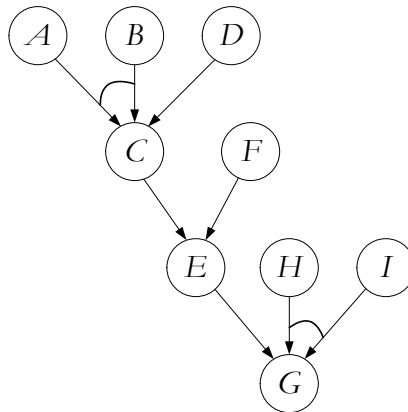


Fig. 5.9: Causal structure generating the coincidence list in table 5.11.

by a causal chain such that AB and D are alternative causes of C , C and F are alternative causes of E , and E and HI are alternative causes of G . In figure 5.9 this structure is graphically depicted.

5.7 Non Interpretable Coincidence Lists

The examples processed by CA thus far illustrated that CA is capable of assigning minimal theories to extended and complex input data. However, the coincidence lists discussed above have been chosen such that a sound causal interpretation has been feasible. Yet, this by far is not generally the case. Consider, for instance, list (a) in table 5.12. Each factor in the frame of that list is a potential effect, i.e. $W = \{A, B, C, D, E\}$, because for none of these factors there is a pair of rows such that a respective factor is the only varying factor as regards that pair. Therefore, none of them would violate (HC) if located in effect position of an underlying causal structure. Moreover, relative to list (a), each of these potential effects is both minimally sufficient and minimally necessary for every other factor in the corresponding frame. Upon processing list (a), CA correctly identifies these minimally necessary conditions of each $Z_i \in W$, but when it comes to framing minimal theories with these conditions CA aborts the analysis of (a). If every factor is minimally sufficient and necessary for every other factor, the direction of underlying causal dependencies is not determinable. In section 3.6.2 we have seen that MT requires causal structures to involve a minimum of two alternative causes for each effect. Otherwise causes and effects are not distinguishable. In fact, this is just what happens in case of list (a). Step 7 does not assign a minimal theory to any of the factors in that list, for none of the effects* is identifiable as a real effect. Or put differently: Relative to the causal background of list (a), CA infers the following dependencies:

					$A \quad B \quad C \quad D \quad E$									
					A	B	C	D	E	A	B	C	D	E
					1	1	1	1	1	1	0	0	0	0
					1	1	1	1	1	0	1	0	0	0
A	B	C	D	E	0	0	1	1	1	0	0	1	0	0
1	1	1	1	1	1	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
(a)					(b)					(c)				

Tab. 5.12: Three coincidence lists that are not causally interpretable.

$$\begin{aligned}
& (B \leftrightarrow A) \wedge (C \leftrightarrow A) \wedge (D \leftrightarrow A) \wedge (E \leftrightarrow A) \wedge \\
& \quad (C \leftrightarrow B) \wedge (D \leftrightarrow B) \wedge (E \leftrightarrow B) \wedge \\
& \quad \quad (D \leftrightarrow C) \wedge (E \leftrightarrow C) \wedge \\
& \quad \quad \quad (E \leftrightarrow D)
\end{aligned} \tag{5.21}$$

No simple minimal theories can be framed with any of these dependencies, for none of the effects* in \mathbf{W} has two alternative minimally sufficient conditions whose disjunction would be minimally necessary for the corresponding effect*. Since no minimal theories are identified in step 7, step 8 does not sanction any causal inferences. Thus, \mathbf{CA} does not assign a causal structure to list (a). Of course, that does not mean that the coincidences in list (a) are not the result of a causal structure after all. \mathbf{CA} 's abstinence from a causal inference merely amounts to stipulating that relative to the presumably homogeneous causal background of list (a), no causal inference is warranted. Had the background been different, however, a causal structure might well have been assignable to the behavior of these factors.

Lists (b) and (c) analogously induce \mathbf{CA} to terminate without drawing causal inferences. As in case of (a), (b) and (c) allow for all factors in $\{A, B, C, D, E\}$ to be effects in accordance with (HC). And again as in case of (a), (b) and (c) do not warrant a determination of the direction of causal dependencies. Relative to list (b), A and B are both minimally sufficient and minimally necessary for each other, while every factor in the triple $\langle C, D, E \rangle$ is minimally sufficient and minimally necessary for the other two factors in the triple. Within the causal background of list (b), thus, the following dependencies hold:

$$(B \leftrightarrow A) \wedge (D \leftrightarrow C) \wedge (E \leftrightarrow C) \wedge (E \leftrightarrow D) \tag{5.22}$$

Finally, upon processing list (c), \mathbf{CA} identifies one minimally sufficient and minimally necessary condition for each of the potential effects in $\{A, B, C, D, E\}$ – \overline{BCDE} for A , \overline{ACDE} for B , \overline{ABDE} for C , \overline{ABCE} for D , and \overline{ABCD} for E . Hence, in list (c) the factors in our exemplary frame behave thus that each is instantiated if and only if all the other factors in the frame are absent. Relative to the causal background of list (c), \mathbf{CA} infers the following dependencies:

$$\begin{aligned}
& (\overline{BCDE} \leftrightarrow A) \wedge (\overline{ACDE} \leftrightarrow B) \wedge (\overline{ABDE} \leftrightarrow C) \wedge \\
& \quad (\overline{ABCE} \leftrightarrow D) \wedge (\overline{ABCD} \leftrightarrow E)
\end{aligned} \tag{5.23}$$

As in case of (5.21), no simple minimal theories can be framed with the dependencies recorded in (5.22) and (5.23), for none of the involved potential effects has two alternative minimally sufficient conditions whose disjunction would be minimally necessary for that potential effect. As a consequence of the failure of step 7 to identify minimal theories for any of the factors in \mathbf{W} , no causal inferences are drawn by step 8. Again however, that does not amount to an output of \mathbf{CA} to the effect that the coincidences in these lists are not the result of a causal structure. \mathbf{CA} 's abstinence from a causal inference merely indicates that relative to the presumably homogeneous causal backgrounds of lists (b) and (c) no causal inferences are warranted.

These examples demonstrate that CA cannot claim to be a *complete* causal inference procedure. It does not assign a causal structure to a coincidence list if and only if the coincidences recorded in that list in fact are generated by a causal structure. As any other procedure of causal reasoning, CA cannot presume the available empirical data to be causally interpretable according to its specific standards. The background of test situations is merely assumed to be homogeneous, which is compatible with unknown causes to be homogeneously present and thus preventing causal inferences. However, CA can claim *correctness* relative to the available data. That means, if CA assigns a causal structure to a coincidence list, the available data in fact warrants a causal inference to just that structure. Or differently put: As the 2^n -method, CA merely infers causal *relevancies*, but is unable to determine causal *irrelevancies*.⁴⁶ Abstinence from a causal inference does not amount to an absence of causal dependencies. However, if CA draws a causal inference, the inferred causal dependencies hold relative to the available data and provided that (HC) and (PEX) are satisfied.

5.8 Empirical Exhaustiveness Violated

As indicated in section 5.5.3, assuming (PEX) amounts to a sweeping idealization of real-life data collection. (PEX) is often violated in scientific practice. Data collection may be incomplete for a host of different reasons. Financial or technical resources may be limited in experimental sciences or nature may fail to provide sufficient data in non-experimental fields. Insufficient data is likely to be one of the main reasons for restricted causal interpretability of that data. Relative to insufficient data, unambiguous and reliable causal reasoning simply is impossible. Nonetheless, even inexhaustive data provides some information as to the underlying causal structure. In order to illustrate this, reconsider the exemplary coincidence lists contained in tables 5.9 and 5.10. While, as shown above, 5.9 is the result of a chain, the structure behind 5.10 is epiphenomenally structured. Notwithstanding these far-reaching causal differences, 5.9 and 5.10 share the four coincidences listed in table 5.13. Against the assumably homogeneous background of the coincidences in table 5.13 *A* and *B* are each minimally sufficient for the other three factors, while the dependencies among *C*, *D*, and *E* are symmetric and, thus, not

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
1	1	1	1	1
1	0	1	1	1
0	1	1	1	1
0	0	0	0	0

Tab. 5.13: The four coincidences shared by lists 5.9 and 5.10.

⁴⁶ Cf. section 5.3.1 above.

causally interpretable. Accordingly, CA assigns the following complex minimal theory to 5.13:

$$(AX_1 \vee BX_2 \Rightarrow C) \wedge (AX_3 \vee BX_4 \Rightarrow D) \wedge (AX_5 \vee BX_6 \Rightarrow E) \quad (5.24)$$

If (PEX) is assumed to be satisfied, (5.24) constitutes CA's final output. Yet, if (PEX) is not taken for granted, subsequent extensions of list 5.13 are possible. Additional coincidences, of course, may drastically change CA's output. Depending on whether 5.13 is complemented in terms of 5.9 or 5.10 CA determines the structure underlying an accordingly complemented list to be the result of a chain and an epiphenomenon, respectively. In both cases, A and B are no longer held to be causally relevant to D . Nonetheless, the causal relevance of A and B to C and E is untouched by extending 5.13 in the sense of either 5.9 or 5.10.

5.13 only features four of the 32 logically possible coincidences over the frame $\{A, B, C, D, E\}$. If (PEX) is not taken for granted, any of the 28 remaining coincidences may be observed later on and integrated into 5.13. However, only a small subset of all these logically possible extensions of 5.13 would be causally interpretable. Suppose, for instance, that all 28 remaining coincidences are in fact incorporated in 5.13. The result is a complete coincidence list, which, as shown in section 5.5.2, does not feature any dependencies among its factors and, thus, is not causally interpretable. The same consequence follows from extending 5.13 in terms of lists (a) or (b) in table 5.14. Both (a) and (b) are not causally interpretable because none of the involved factors can be seen as an effect of an underlying structure. For all factors there is a pair of rows, such that the corresponding factor is the only varying factor in that pair.⁴⁷ Thus, lists (a) and (b) are inhomogeneous coincidence lists and, accordingly, $W = \emptyset$.⁴⁸ An extension of 5.13 as indicated in (c), on the other hand, does not altogether resist a causal interpretation. A and B

A	B	C	D	E	A	B	C	D	E	
1	1	1	1	1	1	1	1	1	1	
+1	1	0	1	1	1	0	1	1	1	A
+1	1	1	0	1	0	1	1	1	1	B
+1	1	1	1	0	+0	0	0	0	1	C
1	0	1	1	1	+0	0	0	1	0	D
0	1	1	1	1	+0	0	1	0	0	E
0	0	0	0	0	0	0	0	0	0	
	(a)					(b)				

Tab. 5.14: (a) and (b) are two extensions of 5.13 that are not causally interpretable. (c) does not allow for an integration of D into the underlying causal structure. Added coincidences are marked by “+”.

⁴⁷ Cf. section 5.5.3.

⁴⁸ Cf. page 232.

are still minimally sufficient for C , D , and E , yet factor D cannot be an effect of the underlying structure any longer. The newly added coincidence features D as only varying factor when compared to the last coincidence listed in (c). Thus, (c) is not homogeneous with respect to D . D cannot be integrated into an underlying causal structure as root factor either, for it is not part of a minimally sufficient condition of any of the possible effects contained in (c). In consequence, CA assigns a minimal theory to (c) that corresponds to (5.24) reduced by the middle conjunct. In the same vein, extensions of 5.13 may be inhomogeneous in regard to any other of the effects in (5.24).

These examples of violated empirical exhaustiveness demonstrate that causal reasoning based on insufficient data is radically underdetermined. Finding A , for instance, to be contained in a minimal theory of D relative to an inexhaustive coincidence list does by no means determine A to be causally relevant to D . Later extensions of such lists may reveal that D is not an effect of the underlying causal structure after all. Minimal theories as (5.13) that are inferred from inexhaustive lists cannot be causally interpreted in a determinate way. Nonetheless, inexhaustive lists allow for at least excluding some causal structures from possibly underlying a respective list. For example, no extension of the list in table 5.13 will ever reveal D to be a possible cause of either A or B . The assumed homogeneity of 5.13 determines that, even though A and B may or may not be contained in an underlying causal structure, *if* they are thus contained, they are *root factors* of that structure. For both A and B there is a pair of rows in 5.13 such that they are the only varying factors in that pair, and, as upon extending coincidence lists no coincidences are removed, all extensions of 5.13 will be inhomogeneous with respect to A and B . Accordingly, CA can be said to identify all causal structures not featuring causal relevance of either C , D , or E to A and B as possibly underlying 5.13. That set of causal structures also includes the empty structure, so to speak, i.e. the structure such that A , B , C , D , and E are mutually causally independent.

Depending on the previous causal knowledge about the structure under investigation the amount of possible extensions of a given coincidence list may be narrowed down significantly. Certain causal reasoning methodologies available in the literature, hence, propose to supplement inexhaustive data by assumptions embedded in the available causal knowledge about the examined process.⁴⁹ Thus, the underdetermination of causal reasoning based on inexhaustive data may be compensated by additional causal assumptions. Or put differently, the amount of elements in the set of structures assigned to an inexhaustive list as 5.13 can be reduced if it is e.g. known beforehand that certain factors cannot be causally related or that some factor can only be the effect and not the cause of some other factor. However, whenever such previous causal knowledge is not available, inexhaustive empirical data inevitably underdetermines causal reasoning.

⁴⁹ Cf. e.g. Ragin (1987), ch. 7, Ragin (2000), pp. 139-141, 198-202, 300-308.

5.9 Summary

This chapter has introduced a procedure of causal reasoning embedded in the regularity theoretic framework developed in chapter 3. Coincidence analysis (CA) differs from existing regularity theoretic inference procedures mainly in two respects: First, CA does not assume test-factors to be causally independent, i.e. test-factors must not be instantiatable in all logically possible combinations in order for corresponding data to be processable by means of CA; second, CA does not, from the outset, presume any factors in analyzed frames to be effects or causes. It has been shown that these two causal assumptions made in the context of the 2^n -method are not indispensable for causal inferences drawn from mere coincidence information. Thus, homogeneity (HC) turns out to be the only causal assumption needed for causal reasoning within a regularity theoretic framework. Instead of resorting to a far-reaching causal assumption as (IND), the ambiguities with respect to assigning causal structures to coincidence information can also be resolved by generating all empirically possible data as regards an investigated causal structure, i.e. by assuming (PEX).

As an immediate consequence thereof, CA is not limited to uncovering causal structures layer by layer. While the 2^n -method is only applicable provided that prior causal knowledge or some fortunate coincidence separates analyzed factor frames in a subset consisting of causally independent test-factors and a subset consisting of a single effect, CA is applicable even without any prior causal knowledge concerning the underlying structure. Apart from generalizing the 2^n -method, CA fills a gap left open by the probabilistic algorithms of causal reasoning as presented in Spirtes, Glymour, and Scheines (2000 (1993)). These algorithms only generate informative outputs provided that analyzed conditional probabilities are lower than 1, i.e. provided that causes do not in a strict sense determine their effects. In contrast, CA is custom-built to deterministic causal dependencies and properly uncovers such dependencies. It is capable of analyzing deterministic causal structures from scratch and in their whole complexity. CA is a *correct* causal inference procedure. It unambiguously assigns a minimal theory and thereby a causal structure to every causally interpretable coincidence list.

APPENDIX

Notation

Events: a, b, c, d, e, e_1, e_2 etc.

(Positive) factors: A, B, C , etc.

Negative factors: $\bar{A}, \bar{B}, \bar{C}$, etc.

Event variables: x, y, z, x_1, x_2 etc.

Factor variables: Z, Z_1, Z_2, Z_3 , etc.

Instantiation: Ae or equivalently $e \in \{x : Ax\}$

Coincidence: $Z_1 Z_2 \dots Z_n =_{df} Z_1 x_1 \wedge Z_2 x_2 \wedge \dots \wedge Z_n x_n$

Coincidence involving negative factors:

$$Z_1 \dots Z_i \bar{Z}_{i+1} \dots \bar{Z}_n =_{df} Z_1 x_1 \wedge \dots \wedge Z_i x_i \wedge K_1 x_1 \dots x_i \wedge \\ \forall x_{i+1} \neg (Z_{i+1} x_{i+1} \wedge K_2 x_1 \dots x_i x_{i+1}) \wedge \dots \wedge \forall x_n \neg (Z_n x_n \wedge K_2 x_1 \dots x_i x_n)$$

(Causally interesting) sufficient condition:

$$Z_1 \mapsto Z_2 =_{df} \forall x (Z_1 x \rightarrow \exists y (Z_2 y \wedge x \neq y \wedge \mathbf{R}xy))$$

Unknown conjuncts: $X_i =_{df} Z_1 x_1 \wedge Z_2 x_2 \wedge Z_3 x_2 \wedge \dots \wedge Z_n x_n, n \geq 1$

Unknown disjuncts: $Y_x =_{df} X_1 \vee X_2 \vee X_3 \vee \dots \vee X_n, n \geq 1$

Double-conditional: $X_1 \vee Y_Z \Rightarrow Z =_{df} (X_1 \vee Y_Z \mapsto Z) \wedge (Z \mapsto X_1 \vee Y_Z)$, such that $X_1 \vee Y_Z$ constitutes a minimally necessary disjunction of minimally sufficient conditions for Z .

Definitions

Factor A (p. 61):

$$A =_{df} \{x : Ax\}, \text{ such that } \{x : Ax\} \neq \emptyset$$

Factor \bar{A} (p. 63):

$$\bar{A} =_{df} \{x : \neg Ax\}, \text{ such that } \{x : \neg Ax\} \neq \emptyset$$

Factor Identity (p. 61):

$$A = B =_{df} \forall x (Ax \equiv Bx)$$

Event Identity (p. 55):

$$e_1 = e_2 =_{df} \tau(e_1) = \tau(e_2) \wedge \sigma(e_1) = \sigma(e_2) \wedge \forall Z(Ze_1 \equiv Ze_2)$$

Instantiation (p. 63):

$$x \text{ instantiates } A \text{ iff } x \in \{x : Ax\}$$

Minimally sufficient condition (p. 96): A conjunction of factors $A_1 \wedge A_2 \wedge \dots \wedge A_n$ is a minimally sufficient condition for a factor B iff

- (a) $A_1 \wedge A_2 \wedge \dots \wedge A_n \mapsto B$
- (b) there is no proper part α of $A_1 \wedge A_2 \wedge \dots \wedge A_n$ such that $\alpha \mapsto B$.

Minimally necessary condition (p. 99): A disjunction of factors $A_1 \vee A_2 \vee \dots \vee A_n$ is a minimally necessary condition for a factor B iff

- (a) $B \mapsto A_1 \vee A_2 \vee \dots \vee A_n$
- (b) there is no proper part β of $A_1 \vee A_2 \vee \dots \vee A_n$ such that $B \mapsto \beta$.

Simple and complex minimal theories (p. 171):

1. A double-conditional with an antecedent consisting of a minimally necessary disjunction of minimally sufficient conditions and a consequent consisting of a single factor is a *simple* minimal theory of that consequent.
2. Every simple minimal theory is a minimal theory.
3. A conjunction of two minimal theories Φ and Ψ is a minimal theory iff
 - (a) at least one factor in Φ is part of Ψ ;
 - (b) Φ and Ψ do not have an identical consequent;
 - (c) for every j , $1 \leq j < n$, in a sequence of entangled factors Z_1, \dots, Z_n , $n \geq 2$: Z_j is contained in the antecedent of the simple minimal theory of Z_{j+1} ;
4. Minimal theories that are not simple are *complex*.

Direct causal relevance (MT_d) (p. 136): A factor A (\bar{A}) is directly causally relevant for a positive factor B iff

- (a) A (\bar{A}) is part of a *simple* minimal theory Φ of B
- (b) A (\bar{A}) stays part of Φ across all extensions of its factor frame.

Indirect causal relevance (MT_i) (p. 136): A factor A (\bar{A}) is indirectly causally relevant for a positive factor B iff there is a sequence S of factors Z_1, Z_2, \dots, Z_n , $n \geq 3$, such that

- (a) $A = Z_1$ ($\bar{A} = Z_1$) and $B = Z_n$
- (b) for each i , $1 \leq i < n$: Z_i is part of the simple minimal theory of Z_{i+1}
- (c) the conjunction of the minimal theories of Z_2, Z_3, \dots, Z_n constitutes a complex minimal theory Ψ
- (d) the factors in Ψ stay part of Ψ across all extensions of Ψ 's factor frame.

Causal relevance (MT) (p. 137): A factor A (\bar{A}) is causally relevant for a positive factor B iff A (\bar{A}) is directly causally relevant for B in terms of MT_d or indirectly causally relevant for B in terms of MT_i .

Causal relevance for negative factors (MT_n) (p. 137): A factor A (\bar{A}) is causally relevant for a negative factor \bar{B} iff the negation of A (\bar{A}) is causally relevant for B according to MT.

Singular causation (SC) (p. 142): An event a is a cause of an event b iff a instantiates a *positive* factor A and b instantiates a factor B , such that

- (a) A is part of a minimal theory Φ of B ,
- (b) Φ is causally interpretable according to MT,
- (c) $a \neq b$ and a and b occur within the same spatiotemporal frame, and
- (d) a is coincident with other events that instantiate a minimally sufficient condition X of B which is part of Φ and contains A .

Homogeneity (HC) (p. 231): The background of m test situations generating a causally analyzed set of coincidences over a factor frame consisting of n factors Z_1, Z_2, \dots, Z_n and containing the set \mathbf{W} of potential effects is causally homogeneous iff for each $Z_i \in \mathbf{W}$ and the set \mathbf{T} of test-factors $\{Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n\}$ the following two conditions hold:

- (1) of all *minimally sufficient conditions* X_k for Z_i (or \bar{Z}_i) at least one conjunct is absent in one of the test situations iff at least one conjunct of X_k is absent in all other $m - 1$ test situations as well, where X_k satisfies the following conditions:
 1. none of Z_1, Z_2, \dots, Z_n and $\bar{Z}_1, \bar{Z}_2, \dots, \bar{Z}_n$ is part of X_k ,
 2. no part of X_k is a genuine test-factor cause or an intermediate factor between a test-factor and Z_i ,
 3. the parts of X_k are *causally relevant* to Z_i (or \bar{Z}_i).
- (2) the rest of a complex cause X_l of Z_i (or \bar{Z}_i) such that $Z_1, \dots, Z_j \in \mathbf{T}$, $j \geq 1$, are part of X_l , does *not vary* in test situations featuring at least one of Z_1, \dots, Z_j .

Proofs

Equivalence of (3.45) and (3.46), p. 121

The logical equivalence of (3.45) and (3.46) follows from the subsequent application of the following arrow definition ($D \rightarrow$):

$$A \rightarrow B \dashv\vdash \neg A \vee B$$

Since expressions within the scope of a quantifier can be treated as propositional expressions, the equivalence of (3.45) and (3.46) can be concisely, yet somewhat informally, proven as follows:

By $D \rightarrow$

$$\begin{aligned} & \forall x_1 \forall x_2 ((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2) \rightarrow \exists y (E y \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \wedge \\ & \forall y (E y \rightarrow \exists x_1 \exists x_2 (((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \end{aligned} \quad (3.45)$$

is equivalent to:

$$\begin{aligned} & \forall x_1 \forall x_2 (\neg((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2)) \vee \exists y (E y \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \wedge \\ & \forall y (\neg E y \vee \exists x_1 \exists x_2 (((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \end{aligned}$$

Then, by changing positions of the disjuncts and, according to $D \rightarrow$, introducing arrows again in both conjuncts one gets:

$$\begin{aligned} & \forall x_1 \forall x_2 (\neg \exists y (E y \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y) \rightarrow \neg((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2))) \wedge \\ & \forall y (\neg \exists x_1 \exists x_2 (((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y) \rightarrow \neg E y), \end{aligned}$$

which by applying quantifier definitions ($\neg \exists \mu A(\mu) \dashv\vdash \forall \mu \neg A(\mu)$) and by switching the order of the conjuncts yields (3.46).

$$\begin{aligned} & \forall y (\forall x_1 \forall x_2 \neg(((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2)) \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y) \rightarrow \neg E y) \wedge \\ & \forall x_1 \forall x_2 (\forall y \neg (E y \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y) \rightarrow \neg((A x_1 \wedge B x_2 \wedge K x_1 x_2) \vee (C x_1 \wedge D x_2 \wedge K x_1 x_2))) \end{aligned} \quad (3.46)$$

(3.50), p. 125

The following is no theorem:

$$\begin{aligned} & (\forall x_1 \forall x_2 ((A x_1 \wedge \neg B x_2 \wedge K x_1 x_2) \rightarrow \exists y (D y \wedge x_1 \neq y \wedge x_2 \neq y \wedge R x_1 x_2 y)) \wedge \\ & \forall x R x x x \wedge \forall x (A x \rightarrow \neg B x)) \vdash \forall x_1 ((A x_1 \rightarrow \exists y (D y \wedge x_1 \neq y \wedge R x_1 x_1 y)). \end{aligned} \quad (.25)$$

This is proven by the following interpretation of (.25):

$$\begin{aligned}
I &= \{c_1, c_2, c_3\} \\
\mathfrak{S}(A) &= \{c_1\} \\
\mathfrak{S}(B) &= \{c_2, c_3\} \\
\mathfrak{S}(D) &= \{c_2\} \\
\mathfrak{S}(K) &= \{(c_1, c_2)\} \\
\mathfrak{S}(\mathbf{R}) &= \{(c_1, c_1, c_1), (c_2, c_2, c_2), (c_3, c_3, c_3), (c_1, c_2, c_3)\}
\end{aligned}$$

Relative to this interpretation, the antecedent of (.25) is true while its consequent turns out false. Therefore, (3.50) is true.

(3.51), p. 125

(3.51) is proven by means of an abbreviated Gentzen/Lemmon-deduction. $\forall\text{E}$ and $\&\text{I}$ are applied in a shortened way.

1	[1]	$\forall x_1 \forall x_2 ((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge \mathbf{R}x_1x_2y)) \wedge \forall x \mathbf{R}xxx \wedge \forall x Kxx \wedge \forall x (Ax \rightarrow \neg Bx)$	A
1	[2]	$(Aa \wedge \neg Ba \wedge Kaa \rightarrow \exists y (Dy \wedge a \neq y \wedge Raay)) \wedge Raaa \wedge Kaa \wedge (Aa \rightarrow \neg Ba)$	1 4 $\times \forall\text{E}$, $A \wedge A \vdash A$
2*	[3]	Aa	A
1	[4]	$Aa \rightarrow \neg Ba$	2 &E
1,2*	[5]	$\neg Ba$	3,4 MPP
1	[6]	Kaa	2 &E
1,2*	[7]	$Aa \wedge \neg Ba \wedge Kaa$	3,5,6 $2 \times \&\text{I}$
1,2*	[8]	$\exists y (Dy \wedge a \neq y \wedge Raay)$	2,7 MPP
1	[9]	$Aa \rightarrow \exists y (Dy \wedge a \neq y \wedge Raay)$	3,8 CP
1	[10]	$\forall x (Ax \rightarrow \exists y (Dy \wedge x \neq y \wedge \mathbf{R}xxy))$	9 $\forall\text{I}$
	[11]	$\forall x_1 \forall x_2 ((Ax_1 \wedge \neg Bx_2 \wedge Kx_1x_2) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge x_2 \neq y \wedge \mathbf{R}x_1x_2y)) \wedge \forall x \mathbf{R}xxx \wedge \forall x Kxx \wedge \forall x (Ax \rightarrow \neg Bx) \rightarrow \forall x (Ax \rightarrow \exists y (Dy \wedge x \neq y \wedge \mathbf{R}xxy))$	1,10 CP

(3.57), p. 131

The following is no theorem:

$$\begin{aligned}
&(\forall x_1 (Ax_1 \wedge \forall x_2 \neg (Bx_2 \wedge Kx_1x_2)) \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge \mathbf{R}x_1y)) \wedge \forall x Kxx \wedge \\
&\forall x \mathbf{R}xx \wedge \forall x (Ax \rightarrow \neg Bx) \vdash \forall x_1 (Ax_1 \rightarrow \exists y (Dy \wedge x_1 \neq y \wedge \mathbf{R}x_1y)).
\end{aligned} \tag{.26}$$

This is proven by the following interpretation of (.26):

$$\begin{aligned}I &= \{c_1, c_2, c_3\} \\ \mathfrak{S}(A) &= \{c_1\} \\ \mathfrak{S}(B) &= \{c_2, c_3\} \\ \mathfrak{S}(D) &= \{c_2\} \\ \mathfrak{S}(K) &= \{(c_1, c_1), (c_2, c_2), (c_3, c_3), (c_1, c_2)\} \\ \mathfrak{S}(R) &= \{(c_1, c_1), (c_2, c_2), (c_3, c_3), (c_1, c_3)\}\end{aligned}$$

Relative to this interpretation, the antecedent of (.26) becomes true while its consequent turns out false. Therefore, (3.57) is true.

INDEX

- A**
- absence 42, 44–45, 118–121
- Albert, David Z. 18, 273
- Anscombe, G. E. M. 31, 273
- antisymmetry 9, 68
- Armstrong, David M. 8, 10, 15, 16, 59, 87, 273
- Aronson, Jerrold. 22, 180, 273
- asymmetry 9, 68, *see* direction of causation
- Austin, J. L. 23, 273
- B**
- Bang-Jensen, Jørgen 73, 137, 273
- Barwise, Jon 29, 31, 273
- Baumgartner, Michael 11, 14, 19, 34, 43, 66, 69, 70, 73, 83, 88, 96, 99, 101, 103–105, 110, 148, 192, 199, 206, 207, 210, 213, 230, 239, 273
- Baylis, C. A. 23, 41, 273
- Belnap, Nuel 19, 273
- Bennett, Jonathan 118, 273
- Blackburn, Simon 32, 279
- Blalock, Hubert M. 197, 274
- Bonissone, Piero P. 148, 279
- Brand, Myles 6, 7, 54, 96, 274, 276
- Broad, Charles D. 7, 88, 90, 96, 119, 274
- Brun, Georg 34, 274
- Butler, R. J. 22, 279
- C**
- Carnap, Rudolf 26, 61, 274
- Cartwright, Nancy 196, 274
- Casati, Roberto . 26, 51, 54, 273, 274
- causal chain 157
 individuation of 168
- causal graph 73–74
 layer 215
 layer of 137, 157
- causal homogeneity *see* homogeneity
- causal irrelevance 83, 201
- causal principles 80–84
 Principle of Causality 81
 Principle of Determinism 18, 80–81
 Principle of Persistent Relevance 83–84
 Principle of Relevance 82–83
- causal reasoning 162, 188
- causal relevance 57, 65–79, 107
 direct 74–75, 107
 for negative factors 123
 indirect 74–75, 107
 non-symmetry of 69
 transitivity of 72
- causal statement 22
- causal structure 66
 minimal complexity of 113
 specification of 115–118
- causation
 general . 3, 9, 49, 56, 66–73, 138
 non-reflexivity of 68
 non-symmetry of 69
 transitivity of 72
 simultaneous 173
 singular 3, 9, 49, 56, 66–73, 138–144
 irreflexivity of 10, 25, 68, 140
 non-symmetry of 69, 140
 non-transitivity of 72, 141
- cause
 alternative 77–78
 common 78–79, 195

complex 75–77
 direct 74
 full 67
 genuine 195
 indirect 74
 chain-graph 137, 157
 chain-problem 160, 243
 Church, Alonzo 28, 274
 Cleland, Carol 274
 co-factor 202, 230
 coincidence 77, 220
 empirically possible 152
 coincidence analysis . . 191, 217–218
 input data 218–228
 coincidence frame 153, 188
 coincidence list 219
 complete 225
 coincidence table 200
 Collins, John . . . 2, 3, 52, 96, 274, 275
 condition 90–101
 minimally necessary . . . 98–101
 minimally sufficient . . . 7, 95–98
 necessary 90–95
 sufficient 90–95
 sufficient 4
 conditional dependence 65
 conditional independence 65
 confounder 230
 Counterfactual Causation 147
 counterfactual equivalence 148
 Cummins, Robert 33, 35, 274
 Cushing, James T. 18, 275
 cycle 79

D

Davidson, Donald 274
 Davis, Wayne A. 19, 68, 274
 df-test 200–202
 digraph 73
 direction of causation . . 9–13, 69, 109,
 111–115
 direction of time 111, 172–174
 double-conditional . . . 103, 135, 171
 Dowe, Phil 21, 180, 181, 274

Ducasse, Curt J. 26, 275

E

Ebbinghaus, Heinz-Dieter . . 139, 275
 edge 73
 Eells, Ellery 10, 70, 72, 275
 effect* 236
 effect field 200
 Ehrling, Douglas 11, 275
 Ellis, Brian D. 118, 119, 275
 energy transfer 180–182
 entangled epiphenomenon 159
 entanglement 159
 causal interpretation of 170
 epi-graph 157
 epiphenomenon . . . 13, 78, 157, 159, 164
 Esfeld, Michael . . . 21, 275, 276, 277,
 279
 event 21–28, 51–56, 113, 125
 negative 45, 128, 141
 event identity 24, 56
 event type *see* factor
 exhaustiveness of coincidence data
 228–230
 explanatory power 168

F

fact 21–28, 127, 128
 general 39
 negative 44
 particular 39
 factor 56–65
 conditionally dependent 65
 conditionally independent . . . 65
 contradictory 65
 contrary 65
 entangled 158, 159, 243
 intermediate 196
 negative . . . 43–45, 63, 121–133
 positive 43, 63
 root . . 73, 79, 191, 207, 216, 233
 subcontrary 65, 109
 factor frame 64
 expansion of 164

persistent 153
 structure conserving expansion
 of 164
 Fair, David 22, 69, 180, 275
 Fetzer, James 19, 68, 274
 ff-test 202–205
 Fine, Kit 26, 275
 Flum, Jörg 139, 275
 four-field test *see* ff-test
 Fraassen, Bas van 18, 275
 Frankel, Lois 69, 275

G

Gödelian equivalents 37
 Gasking, Douglas 10, 114, 275
 Glymour, Clark 18–21, 73, 79,
 81, 148, 192, 215, 223, 257,
 279
 Good, I. J. 18, 275
 Goodman, Nelson 16, 59, 275
 Gottlieb, Dale 33, 35, 274
 graph *see* causal graph
 graph-skeleton 151
 Graßhoff, Gerd 1, 2, 11, 14, 19, 43,
 66, 69, 70, 73, 83, 84, 88,
 96, 99, 101, 103–105, 110,
 111, 148, 192, 199, 206,
 207, 210, 213, 230, 239,
 273, 275
 Gutin, Gregory 73, 137, 273
 Gödel, Kurt 37, 275

H

Hall, Ned 2, 3, 52, 70, 71, 96, 274,
 275
 Halpin, John F. 1, 275
 Hart, Herbert Lionel Adolphus 48,
 119, 120, 275
 Hausman, Daniel 10, 11, 21, 69, 112,
 148, 173, 275, 276
 head 73, 157
 Hempel, Carl Gustav 16, 59, 276
 Hesslow, Germund 72, 276
 Hintikka, Jaakko 36, 276

Hitchcock, Christopher 70, 71, 276
 Hobbes, Thomas 88, 276
 homogeneity 110, 192–205, 230–233
 Honoré, Anthony Maurice 48, 119,
 120, 275
 Horgan, Terence 21, 23, 276
 Huemer, Michael 173, 276
 Hume, David 5, 87, 276

I

independence of test-factors 205–207
 indeterminism 18
 instance 63
 instantiation 57, 63
 intervention 10, 114
 intransitivity 70
 Inwagen, Peter van 173, 274

J

Johnson, W. E. 173, 197, 276

K

K *see* coincidence
 Kant, Immanuel 173, 276
 Kanzian, Christian 23, 51, 276
 Kim, Jaegwon 2, 21, 23, 276
 Kistler, Max 21, 22, 276
 Kitcher, Philip 112, 278
 Kneale, William 8, 277
 Koons, Robert C. 31, 277
 Korb, Kevin B. 81, 277, 279
 Kovitz, Ben 173, 276
 Kvart, Igal 65, 70, 71, 277

L

Lampert, Timm 26–28, 34, 273, 277
 layer *see* causal graph
 Ledwig, Marion 21, 275, 276, 277,
 279
 Lemmon, E. J. 24, 277
 LePore, Ernest 24, 25, 52, 278
 Lewis, David 70, 147, 277
 Lombard, Lawrence Brian 277

M

- Mackie, John L. 11,
13, 14, 21, 22, 28, 33, 43,
46, 67, 88, 96, 99, 104, 105,
113, 119, 239, 277
- Mackie, Penelope 118, 277
- Manchester factory hooters . . 13, 99,
105
- manipulation 10, 114
- Manni, Marco 19, 277
- Marsh, R.C. 26, 39, 44, 279
- Martinich, A. P. 29, 31, 273
- May, Michael 1,
2, 11, 14, 83, 84, 88, 90, 96,
97, 99–101, 104, 109, 111,
192, 202, 206, 207, 210,
213, 275, 277
- McLaughlin, Brian P. 24, 25, 52, 278
- McMullin, Ernan 18, 275
- Meixner, Uwe 25, 277
- Mellor, David Hugh 69, 277, 278
- method of difference 194, 197,
200–202
- Mill, John Stuart 58, 81, 87, 194, 277
- minimal theory 89, 101–106
compatibility with 153
complex 136, 171
part of 103
- Miracchi, Silvano 44, 278
- modularity 213–214
- Molnar, George 8, 277
- Moore, G. E. 26, 277
- mt*-equivalence 153, 157, 186
- multiple minimal theories . . 115–118

N

- Neale, Stephen . . 27, 29, 37, 39, 278
- Nidditch, P. H. 5, 87, 276
- non-symmetry 9, 68
external 10, 112, 114
- non-transitivity 70

O

- Oaklander, L. Nathan 44, 278

- Oppy, Graham 39, 278

P

- Papineau, David 69, 278
- path 73, 195
- Paul, L.A. 2, 3, 52, 96, 274, 275
- Pearl, Judea 148, 151, 278, 279
- Perry, John 29, 31, 273
- Pollock, John 69, 278
- Popper, Karl R. 16, 59, 278
- Price, Huw 114, 278
- Principle of Empirical Exhaustive-
ness 224
strong 228
weak 229
- Prior, Arthur N. 29, 278
- Probabilistic Causation 18, 148
- probabilistic equivalence 148
- progressive localization of a cause 46
- progressive localization of causes 45–
49
- property 25
exemplification 25

Q

- Quine, Willard van Orman . 239, 278
- Quine-McCluskey optimization . 239

R

- R 93, 172, 221
- Ragin, Charles C. 123, 192, 202, 210,
256, 278
- Ramsey, Frank Plumpton 278
- regularity 96
empty 8–9
imperfect 4–6
single-case 15–18
spurious 13–15
- Reichenbach, Hans 23, 41, 278
- relevance *see* causal relevance
- remainder 236
- Rescher, Nicholas 24, 274, 277
- root factor 73
- Rosenberg, Alexander . . . 51, 55, 279
- rule of alternatives 212–213

rule of combination 211–212
 rule of difference 210–211
 Russell, Bertrand 26, 39, 44, 279

S

Salmon, Wesley C. 22, 112, 278, 279
 Sanford, David 11, 112, 279
 Sankey, Howard 81, 119, 273, 277
 Schaffer, Jonathan 119, 120, 181, 279
 Scheines, Richard 18–21, 73, 79,
 81, 148, 192, 215, 223, 257,
 279
 Schilpp, P. A. 37, 275
 Schmidt, Raymund 173, 276
 Schuhmann, Karl 88, 276
 Selby-Bigge, L. A. 5, 87, 276
 self-causation 25, 79
 Simmons, Keith 32, 279
 Simon, Herbert 148, 279
 singular event *see* event
 Slater, John G. 11, 279
 slingshot argument 28
 Sosa, Ernest 21, 186, 276, 277
 spatiotemporal adjacency 174
 spatiotemporal contiguity 174, 177
 spatiotemporal frame 92
 spatiotemporal proximity 92
 specification 46, 49
 by supplementation 116
 causal 117
 Spirtes, Peter 18–21, 73, 79, 81, 148,
 192, 215, 223, 257, 279
 Spohn, Wolfgang . 21, 275, 276, 277,
 279
 Strawson, Peter F. 32, 279
 Suppes, Patrick 10, 18, 21, 22, 69,
 112, 147, 148, 279, 280
 Swain, Marshall 6, 7, 96, 274
 Swartz, Norman 1, 279

T

tail 73, 157, 186
 Taylor, Barry 23, 28, 279
 Taylor, Richard 173, 279

test situation 193, 203, 231
 test-factor 194, 202
 theory of descriptions 37
 Thomas, Wolfgang 139, 275
 Tooley, Michael 21, 186, 276, 277
 trail 138
 transference theory 180
 transitivity 70
 Twardy, Charles R. 81, 279
 2ⁿ-method 210, 214–217

U

Urmson, J. O. 23, 273

V

Varzi, Achille 26, 51, 54, 273, 274
 Vendler, Zeno 22, 279
 Verma, T. S. 148, 279
 vertex 73

W

Walton, D. 54, 276
 Warnock, G. J. 23, 273
 Williamson, John 23, 28, 32, 279
 Wittgenstein, Ludwig 35, 41, 280
 Woodward, James 70, 280
 Wright, Georg Henrik von 10, 114,
 280

X

Xu, Ming 19, 25, 92, 280

Z

Zalta, Edward N. 4, 276

BIBLIOGRAPHY

- Albert, David Z.: *Quantum Mechanics and Experience*, Cambridge: Harvard University Press 1992.
- Anscombe, G. E. M.: Causality and Extensionality, *Journal of Philosophy* 66 (1969), 152–159.
- Armstrong, David M.: *What is a Law of Nature?* Cambridge: Cambridge University Press 1983.
- *A World of States of Affairs*, Cambridge: Cambridge University Press 1997.
- The Open Door: Counterfactual versus Singularist Theories of Causation, in: Sankey, H., editor: *Causation and Laws of Nature*, Dordrecht: Kluwer 1999, 175–186.
- Aronson, Jerrold: On the Grammar of 'Cause', *Synthese* 22 (1971), 414–430.
- Austin, J. L.: Unfair to Facts, in: Urmson, J. O. and Warnock, G. J., editors: *Philosophical Papers*, Oxford: Oxford University Press 1970, 154–174.
- Bang-Jensen, Jørgen and Gutin, Gregory: *Digraphs: Theory, Algorithms and Applications*, London: Springer 2001.
- Barwise, Jon and Perry, John: Semantic Innocence and Uncompromising Situations, in: Martinich, A. P., editor: *The Philosophy of Language*, Oxford: Oxford University Press 1996, 369–381.
- Baumgartner, Michael: Inferring Causal Complexity, (forthcoming).
- and Graßhoff, Gerd: *Kausalität und kausales Schliessen. Eine Einführung mit interaktiven Übungen*, Bern: Bern Studies 2004.
- Baumgartner, Michael and Lampert, Timm: Adequate Formalization, (forthcoming).
- Baylis, C. A.: Facts, Propositions, Exemplification and Truth, *Mind* 57 (1948), 459–479.
- Belnap, Nuel: A Theory of Causation: *Causae Causantes* (Originating Causes) as Inus Conditions in Branching Space-Times, *British Journal for the Philosophy of Science* 56 (2005), 221–253.
- Bennett, Jonathan: Event Causation: The Counterfactual Analysis, *Philosophical Perspectives* 1 (1987), 367–386.
- *Events and Their Names*, Indianapolis: Hackett 1988.
- What Events Are, in: Casati, R. and Varzi, A. C., editors: *Events*, Aldershot: Dartmouth 1996, 137–151.

- Blalock, Hubert M.: *Causal Inferences in Nonexperimental Research*, Chapel Hill: University of North Carolina Press 1961.
- Brand, Myles: Identity Conditions for Events, *American Philosophical Quarterly* 14 (1977), 329–337.
- Simultaneous Causation, in: Inwagen, Peter van, editor: *Time and Cause*, Dordrecht: Reidel 1980, 137–153.
- Brand, Myles and Swain, Marshall: On the Analysis of Causation, *Synthese* 21 (1970), 222–227.
- Broad, Charles D.: The Principles of Demonstrative Induction I-II, *Mind* 39 (1930), 302–317, 426–439.
- Hr. von Wright on the Logic of Induction I-III, *Mind* 53 (1944), 1–24, 97–119, 193–214.
- Brun, Georg: *Die richtige Formel. Philosophische Probleme der logischen Formalisierung*, Frankfurt a.M.: Dr. Hänsel-Hohenhausen 2003.
- Carnap, Rudolf: *Meaning and Necessity*, Chicago: University of Chicago Press 1947.
- Cartwright, Nancy: *How the Laws of Physics Lie*, Oxford: Clarendon Press 1983.
- *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press 1989.
- Causal Diversity and the Markov Condition, *Synthese* 121 (1999), 3–27.
- Casati, Roberto and Varzi, Achille, editors: *Events*, Dartmouth: Dartmouth Publishing Company 1996.
- Church, Alonzo: Review of Carnap's *Introduction to Semantics*, *Philosophical Review* 52 (1943), 298–304.
- Cleland, Carol: On the Individuation of Events, *Synthese* 86 (1991), 229–254.
- Collins, John, Hall, Ned, and Paul, L.A., editors: *Causation and Counterfactuals*, Cambridge: MIT Press 2004.
- Cummins, Robert and Gottlieb, Dale: On an Argument on Truth-Functionality, *American Philosophical Quarterly* 9 (1972), 265–269.
- Davidson, Donald: The Logical Form of Action Sentences, in: Rescher, N., editor: *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press 1967, 81–95.
- *Essays on Actions and Events*, Oxford: Clarendon Press 1980.
- Causal Relations, in: *Essays on Actions and Events*, Oxford: Clarendon Press 1980 (1967), 149–162.
- The Individuation of Events, in: Rescher, Nicholas, editor: *Essays on Actions and Events*, Oxford: Clarendon 1980 (1969), 163–180.
- Davis, Wayne A.: Probabilistic Theories of Causation, in: Fetzer, James, editor: *Probability and Causality*, Dordrecht: D. Reidel 1988, 133–160.
- Dowe, Phil: *Physical Causation*, Cambridge: Cambridge University Press 2000.

- A Counterfactual Theory of Prevention and ‘Causation’ by Omission, *Australasian Journal of Philosophy* 79 (2001), 216–226.
- Ducasse, Curt J.: Propositions, Opinions, Sentences, and Facts, *Journal of Philosophy* 37 (1940), 701–711.
- Ebbinghaus, Heinz-Dieter, Flum, Jörg, and Thomas, Wolfgang: *Einführung in die mathematische Logik*, Heidelberg: Spektrum 1998.
- Eells, Ellery: *Probabilistic Causality*, Cambridge: Cambridge University Press 1991.
- Ehring, Douglas: Causal Asymmetry, *The Journal of Philosophy* 79 (1982), 761–774.
- *Causation and Persistence. A Theory of Causation*, Oxford: Oxford University Press 1997.
- Ellis, Brian D.: *Truth and Objectivity*, Oxford: Basil Blackwell 1990.
- Fair, David: Causation and the Flow of Energy, *Erkenntnis* 14 (1979), 219–250.
- Fine, Kit: First-Order Modal Theories III – Facts, *Synthese* 53 (1982), 43–122.
- Fraassen, Bas van: The Charybdis of Realism. Epistemological Implications of Bell’s Inequality, in: Cushing, James T. and McMullin, Ernan, editors: *Philosophical Consequences of Quantum Theory*, Notre Dame: University of Notre Dame Press 1989a, 97–113.
- *Laws and Symmetry*, Oxford: Clarendon Press 1989b.
- Frankel, Lois: Mutual Causation, Simultaneity and Event Description, *Philosophical Studies* 49 (1986), 361–372.
- Gasking, Douglas: Causation and Recipes, *Mind* 64 (1955), 479–487.
- Gödel, Kurt: Russell’s Mathematical Logic, in: Schilpp, P. A., editor: *The Philosophy of Bertrand Russell*, New York: Tudor 1951, 123–153.
- Good, I. J.: A Causal Calculus I, *British Journal for the Philosophy of Science* 11 (1961), 305–318.
- Goodman, Nelson: *Fact, Fiction, and Forecast*, 4th edition. Cambridge: Harvard University Press 1983.
- Graßhoff, Gerd and May, Michael: Causal Regularities, In Spohn, Ledwig, and Esfeld (2001), 85–114.
- Hall, Ned: Causation and the Price of Transitivity, *Journal of Philosophy* 97 (2000), 198–222.
- Two Concepts of Causation, in: Collins, J., Hall, N., and Paul, L.A., editors: *Counterfactuals and Causation*, Cambridge: MIT Press 2004, 225–276.
- Halpin, John F.: Scientific Law: A Perspectival Account, *Erkenntnis* 58 (2003), 137–168.
- Hart, Herbert Lionel Adolphus and Honoré, Anthony Maurice: *Causation in the Law*, 2nd edition. Oxford: Clarendon Press 1985 (1959).
- Hausman, Daniel: Causation and Experimentation, *American Philosophical Quarterly* 23 (1986), 143–154.

- Hausman, Daniel: *Causal Asymmetries*, Cambridge: Cambridge University Press 1998.
- Hempel, Carl Gustav: *Aspekte wissenschaftlicher Erklärung*, Berlin: de Gruyter 1977 (1965).
- Hesslow, Germund: Two Notes on the Probabilistic Approach to Causality, *Philosophy of Science* 43 (1976), 290–292.
- Hintikka, Jaakko: Identity, Variables and Impredicative Definitions, *Journal of Symbolic Logic* 21 (1956), 225–245.
- Distributive Normal Forms in First-Order Logic, in: *Logic, Language Games and Information*, Oxford: Clarendon Press 1973, 242–286.
- Hitchcock, Christopher: The Intransitivity of Causation Revealed in Equations and Graphs, *Journal of Philosophy* 98 (2001), 273–299.
- Probabilistic Causation, in: Zalta, Edward N., editor: *The Stanford Encyclopedia of Philosophy*, 2002 (URL: <http://plato.stanford.edu/archives/fall2002/entries/causation-probabilistic/>).
- Hobbes, Thomas; Schuhmann, Karl, editor: *Elemente der Philosophie. Erste Abteilung. Der Körper*, Hamburg: Meiner 1997.
- Horgan, Terence: The case against events, *The Philosophical Review* 87 (1978), 28–47.
- Huemer, Michael and Kovitz, Ben: Causation as Simultaneous and Continuous, *The Philosophical Quarterly* 53 (2003), 556–565.
- Hume, David; Selby-Bigge, L. A. and Nidditch, P. H., editors: *A Treatise of Human Nature*, Oxford: Clarendon Press 1978 (1740).
- *An Enquiry Concerning Human Understanding*, Oxford: Oxford University Press 1999 (1748).
- Johnson, W. E.: *Logic in Three Parts*, Volume (I) – (III), New York: Dover Publications 1963 (1924).
- Kant, Immanuel; Schmidt, Raymund, editor: *Kritik der reinen Vernunft*, Hamburg: Meiner 1956.
- Kanzian, Christian: *Ereignisse und andere Partikularien*, Paderborn: Ferdinand Schöningh 2001.
- Kim, Jaegwon: Causation, Nomic Subsumption, and the Concept of Event, *Journal of Philosophy* 70 (1973), 217–236.
- Events as Property Exemplifications, in: Brand, M. and Walton, D., editors: *Action Theory*, Dordrecht: Reidel 1976, 159–177.
- Causes and Events: Mackie on Causation, in: Sosa, Ernest and Tooley, Michael, editors: *Causation*, Oxford: Oxford University Press 1993, 60–74.
- Kistler, Max: Causes as Events and Facts, *Dialectica* 53 (1999), 25–46.
- Causation as Transference and Responsibility, In Spohn, Ledwig, and Esfeld (2001), 115–133.

- Kneale, William: Universality and Necessity, *British Journal for the Philosophy of Science* 12 (1961), 89–102.
- Koons, Robert C.: *Realism Regained : An Exact Theory of Causation, Teleology, and the Mind*, Oxford: Oxford University Press 2000.
- Korb, Kevin B.: Probabilistic Causal Structure, in: Sankey, Howard, editor: *Causation and Laws of Nature*, Dordrecht: Kluwer 1999, 265–311.
- Kvart, Igal: The Counterfactual Analysis of Cause, *Synthese* 127 (2001), 389–427.
- Lampert, Timm: *Wittgensteins Physikalismus*, Paderborn: Mentis 2000.
- Lemmon, E. J.: Comment on “The Logical Form of Action Sentences”, in: Rescher, N., editor: *The Logic of Decision and Action*, Pittsburgh 1967, 96–103.
- *Beginning Logic*, Indianapolis: Hackett 1978 (1965).
- Lewis, David: Causation, *Journal of Philosophy* 70 (1973), 556–567.
- Counterfactual Dependence and Time’s Arrow, *Noûs* 13 (1979), 455–476.
- Postscript to “Causation”, in: *Philosophical Papers*, Volume 2, Oxford: Oxford University Press 1986, 172–213.
- Causation as Influence, *Journal of Philosophy* 97 (2000), 182–197.
- Lombard, Lawrence Brian: Events, *Canadian Journal of Philosophy* 9 (1979), 19–41.
- Mackie, John L.: *The Cement of the Universe. A Study of Causation*, Oxford: Clarendon Press 1974.
- Causes and Conditions, in: Sosa, Ernest and Tooley, Michael, editors: *Causation*, Oxford: Oxford University Press 1993 (1965), 33–55.
- Mackie, Penelope: Causing, Delaying, and Hastening: Do Rains Cause Fires? *Mind* 101 (1992), 483–500.
- Manni, Marco: *Kausalität und Korrelation. Eine Auseinandersetzung mit der Theorie des kausalen Schliessens von Spirtes, Glymour und Scheines*, Bern: Bern Studies forthcoming.
- May, Michael: *Kausales Schliessen. Eine Untersuchung über kausale Erklärungen und Theorienbildung*, Ph.D thesis, Universität Hamburg 1999.
- Meixner, Uwe: *Ereignis und Substanz*, Paderborn: Schöningh 1997.
- Mellor, David Hugh: *Real Time*, Cambridge: Cambridge University Press 1981.
- The Singularity Affecting Facts of Causation, in: *Matters of Metaphysics*, Cambridge: Cambridge University Press 1991 (1987), 201–224.
- *The Facts of Causation*, London: Routledge 1995.
- Transparent and Opaque Causation, in: Spohn, W., Ledwig, M., and Esfeld, M., editors: *Current Issues in Causation*, Paderborn: Mentis 2001, 9–20.
- Mill, John Stuart: *A System of Logic*, London: Longmans 1879 (1843).
- Molnar, George: Kneale’s Argument Revisited, *The Philosophical Review* 78 (1969), 79–89.
- Moore, G. E.: *Some Problems of Philosophy*, London: Allen and Unwin 1953.

- Neale, Stephen: *Facing Facts*, Oxford: Clarendon Press 2001.
- Oaklander, L. Nathan and Miracchi, Silvano: Russell, Negative Facts, and Ontology, *Philosophy of Science* 47 (1980), 434–455.
- Oppy, Graham: Facing Facts? *Australasian Journal of Philosophy* 82 (2004), 621–643.
- Papineau, David: Causal Asymmetry, *British Journal for the Philosophy of Science* 36 (1985), 273–289.
- Pure, Mixed, and Spurious Probabilities and Their Significance for a Reductionist Theory of Causation, in: Kitcher, Philip and Salmon, Wesley C., editors: *Scientific Explanation*, Minneapolis: University of Minnesota Press 1989, 307–348.
- Pearl, Judea: *Causality. Models, Reasoning, and Inference*, Cambridge: Cambridge University Press 2000.
- Pollock, John: *Subjunctive Reasoning*, Dordrecht: Reidel 1976.
- Popper, Karl R.: *Logik der Forschung*, Tübingen: J. C. B. Mohr 1994 (1934).
- Price, Huw: Agency and Causal Asymmetry, *Mind* 101 (1992), 501–520.
- Prior, Arthur N.: Facts, Propositions and Entailment, *Mind* 57 (1948), 62–68.
- Quine, Willard van Orman: The Problem of Simplifying Truth Functions, *The American Mathematical Monthly* 59 (1952), 521–531.
- On Core and Prime Implicants of Truth Functions, *American Mathematical Monthly* 66 (1959), 755–760.
- *Word and Object*, Cambridge: MIT Press 1960.
- On What There Is, in: *From a Logical Point of View*, Cambridge: Harvard University Press 1980a, 1–19.
- Reference and Modality, in: *From a Logical Point of View*, Cambridge: Harvard University Press 1980b, 139–159.
- Two Dogmas of Empiricism, in: *From a Logical Point of View*, Cambridge: Harvard University Press 1980c, 20–46.
- *Methods of Logic*, 4th edition. Cambridge: Harvard University Press 1982.
- Events and Reification, in: LePore, Ernest and McLaughlin, Brian P., editors: *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell 1985, 162–171.
- *Quiddities*, Cambridge: Harvard University Press 1987.
- Ragin, Charles C.: *The Comparative Method*, Berkeley: University of California Press 1987.
- *Fuzzy-Set Social Science*, Chicago: University of Chicago Press 2000.
- Ramsey, Frank Plumpton: Facts and Propositions, in: Mellor, D. H., editor: *Philosophical Papers*, Cambridge: Cambridge University Press 1994 (1927), 34–51.
- Reichenbach, Hans: *Elements of Symbolic Logic*, New York: Collier-Macmillan 1947.

- *The Direction of Time*, Berkeley: University of California Press 1956.
- Rosenberg, Alexander: On Kim's Account of Events and Event-Identity, *The Journal of Philosophy* 71 (1974), 327–336.
- Russell, Bertrand: The Philosophy of Logical Atomism, in: Marsh, R.C., editor: *Logic and Knowledge*, New York: Allen & Unwin 1977, 175–281.
- On the Notion of Cause, in: Slater, John G., editor: *The Collected Papers of Bertrand Russell. Logical and Philosophical Papers 1909–13*, Volume 6, London: Routledge 1992 (1913), 190–210.
- *The Principles of Mathematics*, London: Routledge 1992 (1937).
- Salmon, Wesley C.: An “At-At” Theory of Causal Influence, *Philosophy of Science* 44 (1977), 215–224.
- Sanford, David: The Direction of Causation and the Direction of Conditionship, *The Journal of Philosophy* 73 (1976), 193–207.
- Schaffer, Jonathan: Causation by Disconnection, *Philosophy of Science* 67 (2000), 285–300.
- Simon, Herbert: Spurious Correlation: A Causal Interpretation, *Journal of the American Statistical Association* 49 (1954), 467–479.
- Spirtes, Peter, Glymour, Clark, and Scheines, Richard: *Causation, Prediction, and Search*, 2nd edition. Cambridge: MIT Press 2000 (1993).
- Spohn, Wolfgang, Ledwig, Marion, and Esfeld, Michael, editors: *Current Issues in Causation*, Paderborn: Mentis 2001.
- Strawson, Peter F.: Truth, in: Blackburn, Simon and Simmons, Keith, editors: *Truth*, Oxford: Oxford University Press 1999, 162–182.
- Suppes, Patrick: *A Probabilistic Theory of Causality*, Amsterdam: North Holland 1970.
- Swartz, Norman: *The Concept of Physical Law*, 2nd edition. 2003 (URL: <http://www.sfu.ca/philosophy/physical-law/>).
- Taylor, Barry: *Modes of Occurrence: Verbs, Adverbs, and Events*, Oxford: Blackwell 1985.
- Taylor, Richard: *Action and Purpose*, Englewood Cliffs: Prentice-Hall 1966.
- Twardy, Charles R. and Korb, Kevin B.: A Criterion of Probabilistic Causation, *Philosophy of Science* 71 (2004), 241–262.
- Vendler, Zeno: Effects, Results, and Consequences, in: Butler, R. J., editor: *Analytical Philosophy*, Volume First Series, Oxford: Basil Blackwell 1962, 1–15.
- Causal Relations, *Journal of Philosophy* 64 (1967), 704–713.
- Verma, T. S. and Pearl, Judea: Equivalence and Synthesis of Causal Models, in: Bonissone, Piero P. et al., editors: *Uncertainty in Artificial Intelligence 6*, Amsterdam: Elsevier 1991, 255–268.
- Williamson, John: Facts and Truth, *The Philosophical Quarterly* 26 (1976), 203–216.

- Wittgenstein, Ludwig: *Tractatus logico-philosophicus*, Frankfurt a. M.: Suhrkamp 1995.
- Woodward, James: A Theory of Singular Causal Explanation, *Erkenntnis* 21 (1984), 231–262.
- Wright, Georg Henrik von: *Explanation and Understanding*, Ithaca: Cornell University Press 1971.
- On the Logic and Epistemology of the Causal Relation, in: Suppes, Patrick, editor: *Logic, Methodology, and Philosophy of Science IV*, Amsterdam: North Holland 1973, 293–312.
- *Causality and Determinism*, New York: Columbia University Press 1974.
- Xu, Ming: Causation in Branching Time (I): Transitions, Events and Causes, *Synthese* 112 (1997), 137–192.