Erkenntnis manuscript No.
(will be inserted by the editor)

A Regularity Theoretic Approach to Actual Causation

Michael Baumgartner

Received: date / Accepted: date

Abstract The majority of the currently flourishing theories of actual (token-level) causation are located in a broadly counterfactual framework that draws on structural equations. In order to account for cases of symmetric overdeterminiation and preemption, these theories resort to rather intricate analytical tools, most of all, to what Hitchcock (2001) has labeled *explicitly nonforetracking counterfactuals*. This paper introduces a regularity theoretic approach to actual causation that only employs material (non-modal) conditionals, standard Boolean minimization procedures, and a (non-modal) stability condition that regulates the behavior of causal models under model expansions. Notwithstanding its lightweight analytical toolbox, this regularity theory performs at least as well as the structural equations accounts with their heavy appliances.

Keywords actual causation, regularity theory, overdetermination, preeemption, difference-making

1 Introduction

Theories of token-level or *actual causation* are currently flourishing like hardly ever before. Many of these theories operate within a broadly counterfactual framework that draws on structural equations (cf. e.g. Hitchcock 2001; 2007; Woodward 2003; Halpern and Pearl 2005; Halpern 2008; Halpern and Hitchcock 2010). In order to account for recalcitrant problem cases, such as cases of symmetric overdetermination or preemption, theories employing structural equations resort to rather intricate analytical tools, most of all, to what Hitchcock (2001, 275) has labeled *explicitly nonforetracking counterfactuals*. These nonforetrackers have antecedents in which causes are counterfactually set to non-actual values without their effects changing accordingly. That is, nonforetracking counterfactuals presume counterfactual configurations of causes and their effects that are excluded by the very causal structures under scrutiny. Apart from raising the question to what degree relations of actual causation in the actual world can be clarified by considering non-actual worlds where these relations do not hold, Hall (2007) has pointed out that nonforetrackers create new problem cases for counterfactual accounts, such as cases of switching or short-circuiting (cf. also Hall and Paul 2003).

This paper presents an approach to analyzing actual causation that is located in a broadly regularity theoretic framework. Regularity theorists have repeatedly suggested that their accounts could efficiently capture cases of symmetric overdetermination or preemption (cf. Mackie 1974; Graßhoff and May 2001; Strevens 2007; Baumgartner 2008). However, as the primary target of many regularity theories is causation on the type-level, actual causation is frequently a mere side issue for regularity theorists.[1] Accordingly, rather than developing in detail how their accounts could be adapted to the token-level, they too often content themselves with hinting at the potential of regularity theories of actual causation by means of a few standard examples (e.g. Mackie 1974, 44). This paper intends to make up for the lack of theoretical detail in the regularity theoretic literature on actual causation. It will turn out that the regularity theoretic framework is capable of accounting for structures of overdetermination, preemption, switching, and short-circuiting on the mere basis of material (non-modal) conditionals, standard Boolean minimization procedures, and a (non-modal) permanence or stability condition that regulates the behavior of causal models under model expansions.

The main reason why the vast majority of authors working on actual causation have chosen not to go the regularity theoretic way, of course, is that the standard opinion in the literature has it that regularity theories already fail for their primary analysandum: type causation (cf. e.g. Lewis 1973; Armstrong 1983; Cartwright 1989, 25-29; Hitchcock 2010). In particular, it is claimed that regularity theories cannot distinguish between spurious regularities that hold, for instance, among parallel effects of a common cause and regularities that stem from causal dependencies. While that is indeed the case for Mackie's (1974) well-known INUS-theory or Wright's (1985) NESS-approach, the regularity theoretic literature has, in the meantime, overcome the deficiencies of the INUS- and NESS-theories. Modern regularity theories of type causation, as presented in Graßhoff and May (2001) and Baumgartner (2008) (cf. also Psillos 2009), successfully meet the traditional challenges.

Another reason for the neglect of regularity accounts might be that an intuition apparently shared by many suggests that whether two events are related in terms of actual causation depends on the *intrinsic* properties of the corresponding sequence of events only (cf. e.g. Lewis 1986; Menzies 1996; Hall and Paul 2003). By contrast, a regularity theory entails that whether an event a is an actual cause of another event b, among other things, depends on how a and b relate to other events of the corresponding event types A and B.[2] That is, a regularity theory makes actual causation an extrinsic property of an event sequence. I shall not try to argue over intuitions here. Rather, I will simply introduce the theoretical ease with which a regularity theory handles cases of preemption, overdetermination, switching, short-circuiting and the like, as an incentive to reconsider the intrinsicness intuition.

In the end, this paper's argument in favor of a regularity theoretic approach to actual causation will be of *pragmatic* nature. Glymour et al. (2010) justifiably doubt that, in light of the unmanageable amount of possible counterexamples and of the muddy intuitive background against which theories of actual causation are typically assessed, an entirely satisfactory theory will ever be available. Accordingly, I am not going to claim that a regularity theory is beyond doubt in all conceivable cases. Rather, I am going to argue that it performs at least as well as modern counterfactual accounts. Furthermore, contrary to the latter, a regularity theory achieves its goal by implementing uncontroversial and straightforward conceptual and technical resources.

[1] There are some regularity theoretic proposals that consider token causation to be primary (e.g. Mackie 1965), but the criticism raised against these token-level accounts (e.g. Kim 1971), in my view, shows that these accounts are beyond repair. I shall not pursue the singularist thread in the regularity theoretic literature here.

[2] Hall (2004) shows that counterfactual theories do not respect intrinsicness either (cf. also Maudlin 2004).

Section 2 reviews the basics of a modern regularity theory of type causation and indicates how standard objections can be dealt with. Section 3 then presents the details of a regularity theory of actual causation and illustrates the potential of that theory by applying it to the standard problem cases. Finally, section 4 relativizes the theory to a context-sensitive distinction between typical and atypical scenarios.

2 Regularity theory of type causation

As anticipated above, many regularity theories focus on causation on the type-level as their primary analysandum and take material regularities among event types as their primary analysans.[3] Moreover, regularity theories only aim to analyze *deterministic* causation. The metaphysical question as to the deterministic nature of all causal processes shall be sidestepped here. For our purposes it suffices to note that all causal processes discussed in the structural equations literature on actual causation are explicitly or implicitly assumed to be of deterministic nature, and thus fall into the domain of regularity theories.

To introduce the details of a regularity theory of type causation, some conceptual preliminaries are required. Event types or *factors*, as I call the relata of type causation for short, can be seen as sets of event tokens. If, and only if, a member of such a set occurs, the corresponding factor is said to be *instantiated*. However, not any set of event tokens constitutes a factor that can be involved in causal dependencies. Factors that can be causally related are *suitable* or appropriate for type causation (or for causal modeling), and the factors that can be contained in complex causal structures constitute suitable factor sets. Unfortunately, rendering the relevant notion of suitability precise is a notoriously difficult task, which is often sidestepped in the literature (cf. e.g. Spirtes et al. 2000, 21, 91-92). There exist a few negative suitability standards: for instance, suitable factors do neither correspond to gerrymandered nor gruelike properties (cf. Lewis 1999; Fodor 1997); and different members of a suitable factor set are not related in terms of logical dependence or other forms of dependence that are metaphysically stronger than causation, such as supervenience, constitution, or mereological containment (cf. Hitchcock 2007, 502; Halpern and Hitchcock 2010, Section 4). And there exist some positive suitability standards: for example, suitable factors correspond to (imperfectly) natural properties and all of their instances mutually resemble each other (cf. Lewis 1999), i.e. suitable factors are *similarity sets* of event tokens. Plainly, most of these conditions are vague and only yield suitability by degree.[4] In what follows, the problem of sharpening the relevant suitability standards shall be bracketed. I am simply going to assume that all subsequently employed (sets of) factors meet those standards.

Factors are symbolized by italicized capital letters A, B, C, etc., with placeholders Z_1, Z_2,... representing any factors. Their instances are symbolized by italicized lowercase letters a, b, c, etc., with x, y,... representing any instances. As absences are often causally interpreted as well, factors shall be negatable. The negation of a factor A is written thus: $\overline{A}$. $\overline{A}$ simply represents the absence of an instance of A. Controversial questions as to the

[3] There are some analyses of causation referred to as "regularity theories" that draw on such modal notions as nomic sufficiency (Hausman 1998, 42-43) or counterfactual conditionals (Hall 2004). This terminology, however, blurs the important distinction between empiricist and modal analyses. As this distinction will be of particular importance for this paper, I subsequently reserve the label "regularity theory" for non-modal analyses.

[4] Often the suitability of factors is also rendered dependent on such context-sensitive conditions as salience (Handfield et al. 2008) or farfetchedness (Hitchcock 2001, 287; Woodward 2003, 86–91; Halpern and Pearl 2005, 871). I prefer to first propose a context-independent notion of causation and to postpone all considerations of context-sensitivity to section 4.

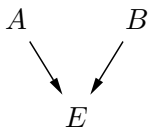


Fig. 1

ontological makeup of the instances of factors or as to what instantiates absences are deliberately ignored in the present context.[5] To avoid these questions the structural equations framework has a very handy terminology on offer: both occurrences and non-occurrences of events are simply understood as random variables taking one of their respective values. Thus, alternatively, factors can be seen as binary variables that take the value 1 whenever a token of the corresponding type occurs and the value 0 whenever no such token occurs.

Clearly, there are certain connections between deterministic causal dependencies and material conditionals. For instance, if it is assumed that factors A and B are the two alternative deterministic causes of E, as depicted in figure 1, it follows that for every instance of type A or B there exists an event of type E and for every event of type E there exists an event of type A or B. Moreover, these A- or B-type events differ from the E-type event (no self-causation) and they occur spatiotemporally proximately or in the same situation (locality). What the relation "x occurs in the same situation as y" amounts to depends on the causal process under investigation and is notoriously vague. For simplicity, I am going to assume that the processes discussed in this paper are sufficiently well known that this relation is properly interpretable.[6] If we introduce the relation Rxy representing "x occurs in the same situation as y", we can express the regularities entailed by the deterministic structure in figure 1 as follows:

$$\begin{aligned} &\forall x((Ax \vee Bx) \rightarrow \exists y(Ey \wedge x \neq y \wedge Rxy)) \wedge \\ &\forall x(Ex \rightarrow \exists y((Ay \vee By) \wedge x \neq y \wedge Rxy)) \end{aligned} \tag{1}$$

Since I shall not be concerned with the requirement as to the non-identity of causes and effects nor with their spatiotemporal proximity, I am going to conveniently abbreviate first-order regularities such as (1) by means of propositional expressions. As a shorthand for (1) I use:

$$A \vee B \leftrightarrow E \tag{2}$$

I take it to be uncontroversial that A and B being the two deterministic causes of E entails that events of type E occur in a situation ω if and only if there is an event of either type A or of type B in ω. Of course, deterministic causal structures in the actual world are not as simple as in figure 1. Single factors do not cause their effects in isolation. Rather, deterministic causes amount to complex conjunctions of co-instantiated factors, i.e. of factors that are instantiated in the same situation and only jointly determine their effect. Moreover, on the type-level, effects can be brought about by several alternative complex causes. That is, regularities entailed by deterministic structures typically are significantly more complex than the one stated in (2). To adequately represent the complexity of regularities induced by

[5] For an interesting suggestion as to how to handle instantiations of absences within an event ontology cf. Handfield et al. (2008, sect. 2.2).

[6] Even though locality is relevant for all theories of causation, it is usually sidestepped in the literature. For more details on the problem of suitably interpreting spatiotemporal proximity for a given causal process cf. Baumgartner (2008).

real-life deterministic structures we thus need to somewhat extend our shorthand notation. I follow Mackie (1974, 66-71) in symbolizing conjunctions of factors by mere concatenation and in introducing variables X_1, X_2,... that stand for open factor conjunctions and variables Y_1, Y_2,... that stand for open disjunctions $X_1 \vee X_2 \vee \ldots \vee X_n$. Furthermore, Mackie (1974, 34-35, 63) relativizes deterministic regularities to what he calls a *causal field*, i.e. to a constant configuration of background conditions. These conventions allow for representing regularities entailed by deterministic structures of greater complexity. A more realistic scenario than the one given in (2) is that A and B are mere parts of alternative causes of E within a field **F**, from which it follows:

$$\text{in } \mathbf{F}: \; AX_1 \vee BX_2 \vee Y_1 \leftrightarrow E \tag{3}$$

In words: in the field **F**, events of type E occur in a situation ω if, and only if, either A is co-instantiated with other factors X_1 in ω or B is co-instantiated with other factors X_2 in ω or further factors Y_1 are instantiated in ω. For brevity, I abstain from making the field-relativity of deterministic regularities explicit in the following. Subsequent regularity statements are, hence, to be understood as implicitly relativized to a given setting of background conditions.

As regularity theorists want to *analyze* deterministic causation in terms of regularities, they not only need a way to infer regularities from deterministic causal structures, but also a way to infer back to causation on the basis of regularities. Contrary to the former direction of entailment, however, the latter is far from straightforward. Most regularities of type (2) or (3) are *spurious* (cf. e.g. Cartwright 1989, 25-29). Therefore, the core task for regularity theorists is to impose constraints on material regularities as (2) and (3) such that the subset of regularities that meet those constraints are those that are non-spurious and, thus, allow for inferring back to causation, i.e. those that are causally interpretable. Modern regularity theories essentially impose two such constraints: (I) causally interpretable material regularities *do not feature redundancies*, and (II) they are *permanent* (or stable). Let us take these constraints in turn.

The most important condition regularities have to satisfy in order to be causally interpretable is what may be called a *principle of non-redundancy*. Causal structures do not feature redundancies. Every cause contained in a type-level causal structure makes a difference to at least one effect in the structure in at least one situation. However, material conditionals—the core analytical tool of regularity theories—are monotonic and, accordingly, tend to feature a host of redundancies. If AB is sufficient for E, so is ABZ (i.e. $AB \rightarrow E \;\vdash\; ABZ \rightarrow E$), and if $A \vee B$ is necessary for E, so is $A \vee B \vee Z$ (i.e. $E \rightarrow A \vee B \;\vdash\; E \rightarrow A \vee B \vee Z$). In both cases, Z may be interpreted to stand for any arbitrary factor. That means sufficient and necessary conditions can only be causally interpreted if all redundancies are removed from them, i.e. if they are rigorously minimized. To this end, modern regularity theories draw on the notions of a minimally sufficient condition and of a minimally necessary condition (cf. Graßhoff and May 2001; Baumgartner 2008). AX_1 is a *minimally sufficient condition* of E iff $AX_1 \rightarrow E$ and for no proper part α of AX_1: $\alpha \rightarrow E$, where a proper part of a conjunction is that conjunction reduced by at least one conjunct. $AX_1 \vee BX_2 \vee Y$ is a *minimally necessary condition* of E iff $E \rightarrow AX_1 \vee BX_2 \vee Y$ and for no proper part β of $AX_1 \vee BX_2 \vee Y$: $E \rightarrow \beta$, where a proper part of a disjunction is that disjunction reduced by at least one disjunct.

Minimizing sufficient and necessary conditions amounts to systematically testing whether they contain sufficient and necessary proper parts and to eliminating redundant parts. Such systematic redundancy testing requires sufficient and necessary conditions to be

given in a particular syntactic form: disjunctive normal form.[7] To have a handy label for the resulting minimally necessary disjunctions of minimally sufficient conditions, I (following Graßhoff and May 2001) introduce the notion of a *minimal theory*.

Minimal Theory: A minimal theory Φ of a factor E is a minimally necessary disjunction of minimally sufficient conditions (in disjunctive normal form) of E, such that (i) the conjuncts in each disjunct of Φ are instantiated in the same situation, (ii) E is instantiated in the same situation as its minimally sufficient conditions, and (iii) the instances of E differ from the instances of its minimally sufficient conditions.

To illustrate, reconsider the simple structure depicted in figure 1. In this structure, A and B each are sufficient for E and, as they do not contain proper parts, they do not contain sufficient proper parts. Hence, A and B each are minimally sufficient for E. The disjunction $A \vee B$, in turn, is necessary for E and neither of its proper parts is itself necessary for E, for according to the structure in figure 1, A and B are alternative causes of E, which is only the case if neither A nor B is redundant to account for all instances of E. That is, there are circumstances such that A makes a difference to E independently of B, and vice versa. It hence follows that there exist instances of E without instances of A, i.e. instances of E that are caused by instances of B only, and there exist instances of E without instances of B, i.e. instances of E that are caused by instances of A only.[8] Overall, the structure in figure 1 not only entails (1), but moreover (4):

$$\begin{aligned} &\forall x((Ax \vee Bx) \rightarrow \exists y(Ey \wedge x \neq y \wedge Rxy)) \wedge \\ &\forall x(Ex \rightarrow \exists y((Ay \vee By) \wedge x \neq y \wedge Rxy)) \wedge \\ &\neg\forall x(Ex \rightarrow \exists y(Ay \wedge x \neq y \wedge Rxy)) \wedge \\ &\neg\forall x(Ex \rightarrow \exists y(By \wedge x \neq y \wedge Rxy)) \end{aligned} \tag{4}$$

To suitably abbreviate the formal expression of minimal theories in our shorthand notation, I introduce the operator "$\Rightarrow$", which does not only state regularities among factors as expressed in (1) but moreover determines sufficient and necessary conditions to be *minimal*. This allows for abbreviating (4) in terms of (5):

$$A \vee B \Rightarrow E \tag{5}$$

(5) is the minimal theory *over* the set $\{A, B, E\}$ expressing the minimized deterministic dependencies regulating the behavior of E as induced by the deterministic structure in figure 1. $A \vee B$ is the antecedent of the minimal theory (5) and E its consequent. A factor Z is said to *be part of* a minimal theory Φ of E iff Z is a conjunct of at least one disjunct in the antecedent of Φ.

The notion of a minimal theory takes us a long way towards identifying the subset of material regularities that allow for inferring back to causation, for it turns out that minimal theories do not state spurious regularities. That is, if the causally interpretable regularities are restricted to minimal theories, spurious regularities are precluded from a causal interpretation. To see this, consider the deterministic structure depicted in figure 2. In this structure,

[7] There exist several Boolean procedures that algorithmically minimize sufficient and necessary conditions, the most well known being Quine-McCluskey optimization (cf. Quine 1959). For an alternative cf. Baumgartner (2009).

[8] Plainly, the non-redundancy principle does not require relevant difference-making circumstances to exist in the past or the present of a particular causal analysis. These circumstances simply need to exist in a tenseless sense (in the domain of quantification).

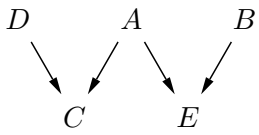


Fig. 2

C and E are two parallel effects of the common cause A. In addition, there exists one further alternative cause for C and E each: D for C and B for E. Even though this structure is again artificially simple, it suffices for our current purposes, for it yields spurious regularities. For instance, it entails that C in combination with the absence of D, i.e. $C\overline{D}$, is minimally sufficient for E without $C\overline{D}$ being a complex cause of E. Whenever $C\overline{D}$ is instantiated, A is instantiated as well, for no effect occurs without any of its causes. Hence, if D is absent, A must be present to account for C. Furthermore, since A determines E in structure 2, it follows that $C\overline{D}$ is sufficient for E as well. Of course, $C\overline{D}$ is moreover part of a necessary condition of E:

$$C\overline{D} \vee A \vee B \leftrightarrow E \tag{6}$$

As $C\overline{D}$ is composed of *INUS-conditions* of E (cf. Mackie 1974, 62), Mackie's INUS-theoretical variant of a regularity theory is forced to interpret $C\overline{D}$ as complex cause of E, which, according to the structure in figure 2, is false. Structures as this one are ubiquitous—the most famous concrete example being the so-called *Manchester Factory Hooters* example, in light of which Mackie (1974, 83-87) ultimately abandoned the attempt to provide a genuine regularity theoretic analysis of type causation. However, (6) is not a minimal theory of E, for $C\overline{D} \vee A \vee B$ is only necessary but not minimally necessary for E. It contains one necessary proper part: $A \vee B$. Whenever E occurs, A or B occur as well. The left-hand side of (6) has no other necessary proper part. $C\overline{D} \vee A$ is not necessary for E, because according to figure 2 E may occur without $C\overline{D}$ and A—say, when CD is given along with $\overline{A}$ and B. Neither is $C\overline{D} \vee B$ necessary for E: E may occur without $C\overline{D}$ and B—for example, when CD is given in combination with $\overline{B}$ and A. Among the elements of the necessary condition of E in (6) the following asymmetry holds, which allows for eliminating $C\overline{D}$: $C\overline{D}$ is sufficient for $A \vee B$, while $A \vee B$ is not sufficient for $C\overline{D}$. That means, while $A \vee B$ makes a difference to E independently of $C\overline{D}$, the converse does not hold. The minimal theory of E entailed by figure 2 is not (6) but (5).

These considerations reveal the principal deficiency of Mackie's INUS-theory and Wright's NESS-account. Both of these theories do not minimize material regularities rigorously enough. Mackie and Wright only call for a minimization of sufficient conditions. Yet, necessary conditions may contain redundancies as well.[9] By contrast, causal structures do not feature any redundancies whatsoever. By rigorously minimizing both sufficient and necessary conditions those factors are filtered out that under some circumstances actually make a difference to the outcome. Thereby it becomes possible to distinguish between regularities that stem from causal dependencies and regularities that are spurious.

Minimizing necessary conditions also prevents a causal interpretation of so-called *accidental regularities*, e.g. of regularities that exist because involved factors have only very few instances that, by chance, happen to coincide with specific effects (cf. Armstrong 1983, 15-17). To illustrate, assume that Harold Bride, the junior wireless operator on the Titanic,

[9] For this reason, recent attempts to reanimate Wright's NESS-test for the analysis of actual causation, as can be found in Baldwin and Neufeld (2004) or Halpern (2008), are bound to fail.

for the first (and only) time in his life lit a Havana cigar moments before the ship hit the iceberg. Suppose, moreover, that we define a factor H that has Harold's lighting of a cigar as its only instance. Then, if we let W stand for the occurrence of a shipwreck, the conditional $H \rightarrow W$ is true and, moreover, the instances of its antecedent and consequent differ and are spatiotemporally proximate. As H does not comprise proper parts, it is not only sufficient, but also minimally sufficient for W. H is not the only minimally sufficient condition of W. Shipwrecks are regularly preceded by storms (S) or fires (F) or collisions with icebergs (I) etc. The particular instance of W constituted by the sinking of the Titanic was preceded by an instance of I. Nonetheless, there is a necessary condition of W that contains H, *viz.* $H \vee S \vee F \vee I \vee Y_1$. Yet, that condition, analogously to the necessary condition of E given in (6), is not minimal, for it holds that $H \rightarrow I$ and $\neg(I \rightarrow H)$. Hence, H makes no difference to E independently of I and is therefore redundant.

Causal models are always relativized to the set of factors considered. This relativization is of particular relevance to proper minimizations of sufficient and necessary conditions, for the elimination of all redundancies essentially hinges on the diversity of that factor set. In contexts of epistemic limitation, notably in contexts of causal discovery, the factor set of an analysis may well not be diverse enough to allow for complete minimizations. Therefore, material regularities that are maximally minimized relative to such a context cannot be unconditionally interpreted causally. As indicated above, the non-redundancy requirement (I) is not sufficient to guarantee the causal interpretability of material regularities in all circumstances. We additionally need to impose a *permanence* constraint (II).

What that supplementary constraint amounts to can again be illustrated by means of the structure in figure 2. Suppose the scientific discipline investigating the causal structure depicted in figure 2 starts by considering the factors in the set $\mathcal{F}_1 = \{B, C, D, E\}$. Relative to $\mathcal{F}_1$ it is discovered that $C\overline{D}$ and B are each minimally sufficient for E. At the same time, the scientists investigating the behavior of E are confronted with instances of E in situations where both $C\overline{D}$ and B are absent. That is, the set $\mathcal{F}_1$ does not feature a necessary condition of E. In consequence, the researchers infer the existence of further unmeasured causes of E outside of $\mathcal{F}_1$. They hence conjecture the validity of the following minimal theory (with Y_1 running over the unmeasured causes):

$$C\overline{D} \vee B \vee Y_1 \Rightarrow E \tag{7}$$

Now suppose that after a while of further investigation the initial set $\mathcal{F}_1$ is expanded to $\mathcal{F}_2 = \{A, B, C, D, E\}$. Relative to $\mathcal{F}_2$, it is then discovered that the formerly unmeasured factor A constitutes an additional minimally sufficient condition of E. Moreover, now the scientists can account for all instances of E: whenever E is instantiated, there is an instance of $C\overline{D}$ or A or B. Thus, a necessary condition of E has been discovered. This finding raises the question whether that necessary condition is minimal. As we have seen above, that is not the case. The discovery of A renders $C\overline{D}$ redundant, which, accordingly, drops out of a minimized necessary condition. That $C\overline{D}$ appeared to make a difference to E turns out to have been a mere by-product of the limited diversity of $\mathcal{F}_1$. Expanding $\mathcal{F}_1$ to $\mathcal{F}_2$ reveals that the regularity $C\overline{D} \rightarrow E$ is spurious. Accordingly, $C\overline{D}$ is no longer part of a minimal theory of E over $\mathcal{F}_2$.

In order to reveal the spuriousness of regularities and the corresponding redundancy of elements of minimal theories, expansions of factor sets must be suitable. A *suitable expansion* $\mathcal{F}_j$ of a factor set $\mathcal{F}_i$ is a superset of $\mathcal{F}_i$, i.e. $\mathcal{F}_i \subseteq \mathcal{F}_j$, which is the result of introducing factors into $\mathcal{F}_i$ that are all suitable for causal modeling and that are logically independent of the elements of $\mathcal{F}_i$ and do not introduce relationships of supervenience, of constitution,

or of mereological containment. A suitable expansion $\mathcal{F}_j$ of $\mathcal{F}_i$ reveals that a factor $Z_i \in \mathcal{F}_i$ which is part of a minimal theory Φ_i of Z_n over $\mathcal{F}_i$ is redundant to account for an effect Z_n if, and only if, Z_i is not part of a minimal theory Φ_j of Z_n over $\mathcal{F}_j$. If there does not exist a suitable expansion $\mathcal{F}_j$ that reveals the redundancy of Z_i, I shall say that Z_i is *permanently non-redundant* for Z_n. That is, a material regularity $Z_i \rightarrow Z_n$ is causally interpretable only if it permanently satisfies the non-redundancy principle, i.e. if Z_i is permanently non-redundant for Z_n. More generally, a minimal theory Φ_i of a factor Z_n over a factor set $\mathcal{F}_i$ is causally interpretable only if for all suitable expansions $\mathcal{F}_j$ of $\mathcal{F}_i$ there exists a minimal theory Φ_j of Z_n over $\mathcal{F}_j$ such that all factors Z_i that are part of Φ_i are also part of Φ_j. Or inversely put: a minimal theory is causally interpretable only if none of its members are rendered redundant by suitably expanding the corresponding factor set.[10]

It must be emphasized that eliminating spurious regularities by systematically extending analyzed factor sets and rigorously removing redundancies presupposes that causal structures are of a certain *minimal complexity*. Take an almost empty universe that only comprises events of types A, C, and E such that these factors correspond to fundamental properties, i.e. properties that cannot be further analyzed. Moreover, assume, for the sake of the argument, that A is a common cause of C and E. It follows that any of those three factors is instantiated if, and only if, any other of those factors is instantiated. A, C, and E are mutually biconditionally dependent. As no other factors are involved in this structure, it cannot be extended; nor can it be modeled on a more fine-grained level to enhance complexity via specification. Hence, the dependency between C and E is spurious and both free of redundancies and permanent. Indeed, the most well-known counterexamples to regularity theories are exactly of this simplistic form. Figure 2, however, shows that if there exists only one further alternative cause for each of C and E, dependencies among C and E are rendered redundant and, thus, identifiable as spurious on regularity theoretic grounds. Accordingly, deterministic structures that are amenable to a regularity theoretic treatment must feature at least two alternative causes for each effect. This does not mean that analyzed factor sets must actually include two alternative causes for each effect; it just means that factor sets must be extendable to include two causes. In light of the enormous causal complexity of the world we live in, it is fair to assume that all type-level causal structures de facto exhibit that minimal complexity. In fact, I would want to claim that determinate causal dependencies only exist in complex worlds, for permanent biconditional dependencies among three factors, in principle, cannot be unambiguously interpreted causally. Yet, even if somebody wants to insist that toy worlds, as the one described above, may feature causal dependencies, these worlds do not show that regularity theories fail to adequately account for type causation as found in the actual world. At best, these simplistic toy examples indicate that regularity theories are not conceived for toy worlds.

With this caveat in mind, I propose the following analysis of type causation (in the actual world):

Type causation (TC): A factor A is a type-level cause of a factor E iff there exists a factor set $\mathcal{F}_i$, where $A, E \in \mathcal{F}_i$, such that (i) A is part of a minimal theory Φ_i of E over $\mathcal{F}_i$, and (ii) for all suitable expansions $\mathcal{F}_j$ of $\mathcal{F}_i$, there exists a minimal theory Φ_j of E over $\mathcal{F}_j$ such that A is part of Φ_j—in short, iff A is permanently non-redundant for E.

[10] Due to the universal (or negative existential) nature of this permanence requirement its satisfaction may be difficult to establish in contexts of epistemic limitations. Plainly though, such uncertainties are a trademark problem encountered in contexts of causal discovery. May (1999, 74) has shown that spurious regularities have certain features that allow for their identification even prior to complete expansions of corresponding factor sets. Halpern and Hitchcock (2010) have recently emphasized that acquiring structural stability across expansions of causal models is of utmost importance for the structural equations framework as well.

For simplicity of exposition and as my primary concern here is not with contexts of causal discovery, regularities expressed by a minimal theory Φ_i over a set $\mathcal{F}_i$ shall be assumed to satisfy the permanence constraint (TC.ii) by default in the following.

Before we turn to actual causation, let me emphasize a few features of (TC) that will be important for the ensuing discussion of actual causation. First, contrary to what critics of regularity theories have often claimed (cf. e.g. Armstrong 1983, ch. 2), material regularities may allow for distinguishing between causes and effects, i.e. for identifying the direction of causation. To see this, reconsider the minimal theory (5) which exhibits the alternative causes of E in the structure of figure 2. The regularities among E and its alternative causes A and B that are entailed by this structure exhibit an important non-symmetry: A and B each determine E, but E does neither determine A nor B, but only the disjunction $A \vee B$. A complete instantiation of a (complex) cause determines the corresponding effect factor, but the latter does not determine which of its alternative type-level causes is responsible for its instantiation in a particular situation. This is the non-symmetry of determination (cf. Baumgartner 2008).[11] Distinguishing between causes and effects on the basis of this non-symmetry, of course, presupposes that an analyzed factor set comprises at least two complete causes for a corresponding effect.[12] If a factor Z_1 is both necessary and sufficient for another factor Z_2 relative to a given factor set $\mathcal{F}_i$, i.e. if it holds in $\mathcal{F}_i$ that $Z_1 \leftrightarrow Z_2$, conditional dependencies are symmetric and do not permit a distinction between cause and effect. In that case, identifying one of Z_1 and Z_2 as effect (or cause) requires either imposing some external non-symmetry, as most commonly the direction of time, or extending $\mathcal{F}_i$ until another condition is found that is minimally sufficient for one of Z_1 and Z_2, and independent of the other.

Second, by conjunctively concatenating minimal theories, causal structures of arbitrary complexity can be represented on regularity theoretic grounds. For instance, (8) represents a causal chain such that A and B are causes of C which, in turn, is a cause of E; or (9) exhibits a common cause structure in which B is a common cause of C and E:

$$(AX_1 \vee BX_2 \vee Y_1 \Rightarrow C) \wedge (CX_3 \vee DX_4 \vee Y_2 \Rightarrow E) \tag{8}$$

$$(AX_1 \vee BX_2 \vee Y_1 \Rightarrow C) \wedge (BX_3 \vee DX_4 \vee Y_2 \Rightarrow E). \tag{9}$$

Third, over one set $\mathcal{F}_i$, there may exist several minimal theories for one effect. To see this, consider the neuron diagram in figure 3. As such diagrams are omnipresent in the actual

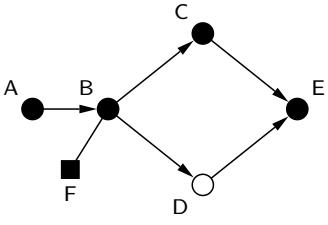


Fig. 3

[11] One might be inclined to argue that some causes may also have alternative effects and that, in such cases, the direction of determination is reversed. However, note that causes that bring about one effect in one situation and another effect in another situation are not deterministic. In deterministic structures, which constitute the domain of regularity theories, there are no causes with alternative effects.

[12] Similarly, to orient edges in causal Bayes nets at least two alternative paths are required that have a common end node, so-called *unshielded colliders* (cf. especially Pearl 2000, 51-57).

causation literature (cf. Collins et al. 2004; Hall 2007; Hitchcock 2009), their graphical features do not need explaining. Suffice it to say that, contrary to the graphs in figures 1 and 2, a neuron diagram does not represent a type-level but a token-level structure. The diagram in figure 3 exhibits a switching process where the firing of neuron A triggers B to fire, which in combination with a firing of the switch F stimulates C and, finally, E. Still though, this token-level process instantiates an underlying type-level structure, which e.g. rules that in situations where F does not fire the stimulatory influence of A on E is transmitted via D. If we model this underlying type-level structure relative to the factor set $\mathcal{F}_3 = \{A, B, C, D, E, F\}$, where each element simply represents the firing of the respective neuron, we find four minimally sufficient conditions for E: A, B, C, D.[13] A disjunctive concatenation yields a necessary condition of E: if E fires there is also firing of A or B or C or D. Overall, it holds:

$$A \vee B \vee C \vee D \leftrightarrow E \tag{10}$$

(10) is not a minimal theory because the necessary condition on its left-hand side contains three proper parts that are themselves necessary for E:

$$A \Rightarrow E \;\; ; \;\; B \Rightarrow E \;\; ; \;\; C \vee D \Rightarrow E. \tag{11}$$

That is, the type-level structure of which the process in figure 3 is an instance yields three different minimal theories for E over $\mathcal{F}_3$.

In general, the behavior of an outcome of a deterministic structure can be expressed as a function of its direct causes or of indirect causes further up on a causal chain. Of course, the mere regularities stated by the minimal theories in (11) are not sufficient to determine which of these theories exhibits the direct causes of E. Moreover, the regularities in $A \Rightarrow E$ and $B \Rightarrow E$ do not even distinguish between causes and effects. But if we assume that we have some way of orienting these dependencies, say, because figure 3 only depicts a substructure of a more complex neuron diagram that features at least two alternative causes for each effect or because we impose a temporal ordering on the firings of these neurons, the minimal theories in (11) can be oriented and grouped into *direct* and *indirect* theories of E over $\mathcal{F}_3$: $C \vee D \Rightarrow E$ is the direct minimal theory, and $A \Rightarrow E$ and $B \Rightarrow E$ are indirect theories. Similarly, there is one direct and one indirect minimal theory for each of C and D: $BF \Rightarrow C$, $AF \Rightarrow C$, $B\overline{F} \Rightarrow D$, $A\overline{F} \Rightarrow D$.

A type-level structure is completely characterized by direct minimal theories only. The indirect theories are mere logical consequences of a complete characterization on the basis of direct theories. Nonetheless, as will become apparent shortly, indirect minimal theories are important to evaluate the non-redundancy of factors for effects further down on a causal chain. For transparency, I subsequently label indirect theories with an index: $(X_1 \Rightarrow Z_1)_i$. In sum, the complex minimal theory over $\mathcal{F}_3$ representing the type-level structure underlying the neuron diagram in figure 3 is this:

$$\begin{gathered}(A \Rightarrow B) \wedge (BF \Rightarrow C) \wedge (B\overline{F} \Rightarrow D) \wedge (C \vee D \Rightarrow E) \wedge \\ (AF \Rightarrow C)_i \wedge (A\overline{F} \Rightarrow D)_i \wedge (A \Rightarrow E)_i \wedge (B \Rightarrow E)_i\end{gathered} \tag{12}$$

Finally, it is important to note that (TC) yields a *non-transitive* notion of type causation. Factors may make a difference to their direct effects and no difference to effects that are located further down on a causal chain. A factor Z_1 may be part of a minimally sufficient condition of Z_2 which, in turn, is part of a minimally sufficient condition of Z_3, without

[13] F is not part of a minimally sufficient condition of E because the firing of the switch F can be eliminated from every sufficient condition without sufficiency for E being lost.

Z_1 being contained in a minimally sufficient condition of Z_3. The structure characterized in (12) provides an example. In this structure, F makes a difference to whether the stimulatory impulse is transmitted via instances of C or D, but E is instantiated independently of the way of transmission. Correspondingly, F is part of a minimal theory of C and C is part of minimal theory of E, but F is not part of a minimal theory of E. That is, according to (TC), F is a cause of C and C is a cause of E, but F is no cause of E. The non-transitivity of (TC) is the reason why indirect minimal theories are needed to assess non-redundancies in difference-making along causal paths.

3 Regularity theory of actual causation

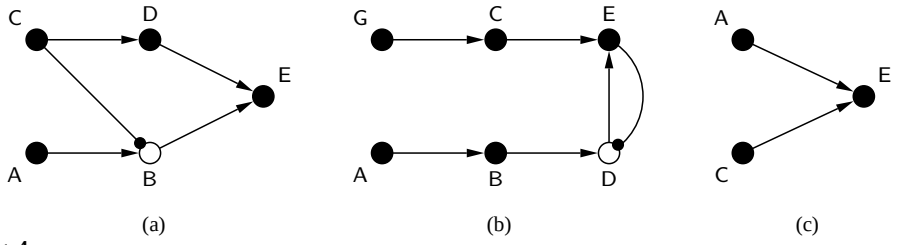
Let us now apply that type-level theory to analyzing actual causation. To avoid intricate questions regarding the ontological makeup of the relata of actual causation, I shall simply use the neutral term 'token' to refer to the relata of actual causation. The basic idea behind a regularity theoretic analysis of actual causation can then be stated very simply: two tokens are causally related if, and only if, they properly instantiate an underlying type-level structure. Actual causation, hence, is a secondary relation that hinges on how corresponding tokens relate to other tokens of the same types and on how these types relate to each other. To spell this basic idea out in more detail, we first have to clarify what it means for two tokens to *properly instantiate* a type-level structure. To this end, one auxiliary notion is required that I borrow from Hitchcock (2001) and adapt to the regularity theoretic context: the notion of an *active causal route*. Roughly, an active causal route is a causal path of a type-level structure that is instantiated in a concrete situation. More specifically:

Active causal route: Relative to a factor set $\mathcal{F}_i$, Z_1 is connected to Z_n by an active causal route in a concrete situation ω iff there is a sequence $\langle Z_1, \ldots, Z_n \rangle$ in $\mathcal{F}_i$ such that for each i, $1 \leq i < n$: (i) Z_i is contained in a direct minimal theory Φ_{i+1} of Z_{i+1} over $\mathcal{F}_i$; and (ii) in ω, Z_i is co-instantiated with all factors X_i constituting a minimally sufficient condition Z_iX_i of Z_{i+1} in Φ_{i+1}.

To illustrate, reconsider the scenario depicted in figure 3 along with the corresponding complex minimal theory (12) over $\mathcal{F}_3$. In (12), A is part of a direct theory of B which is contained in a direct theory of C which, in turn, is part of a direct theory of E. Moreover, in diagram 3, A is co-instantiated with all other factors constituting a minimally sufficient condition AX_1 of B—in this case, of course, X_1 amounts to the empty set because A is itself minimally sufficient for E—, B is co-instantiated with F, and C, which is itself minimally sufficient for E, is instantiated as well. Hence, in that neuron firing process, A is connected to E by an active causal route. Note that the notion of an active route is relativized to a factor set. As a consequence, two factors may be connected by an active route relative to one factor set, but not relative to another. Overall, that two tokens properly instantiate an underlying type-level structure means that those tokens connect two corresponding types by an active causal route relative to a set $\mathcal{F}_i$ and relative to all suitable expansions of $\mathcal{F}_i$, i.e. that those tokens *permanently* connect the corresponding types by an active route.

The notion of an active causal route now enables us to define actual causation on the basis of (TC):

Actual causation (AC): A token a is an actual cause of a different token e iff there exists a factor set $\mathcal{F}_i$ that contains two factors A and E such that (i) A is part of minimal theory Φ_i of E over $\mathcal{F}_i$ and A is permanently non-redundant for E, i.e. A is a type-level cause of E according to (TC); and (ii) for all suitable expansions $\mathcal{F}_j$ of $\mathcal{F}_i$ (which include



(a) (b) (c)

Fig. 4

$\mathcal{F}_i$ itself), A is on an active causal route to E relative to $\mathcal{F}_j$ such that a and e are the instances of A and E, respectively, on this route.

As in case of (TC.ii), the permanence constraint (AC.ii) is of universal form and may, hence, be difficult to establish in contexts of causal discovery. In order not to get entangled in intricate questions of causal discovery, I shall simply assume that—if not explicitly stated otherwise—the subsequently discussed neuron diagrams *completely* represent corresponding causal processes, i.e. no further causes or causal intermediaries are left out. This greatly limits the extendability of relevant factor sets and yields that (AC.ii) can be visibly (in)validated.

The remainder of this section demonstrates the potential of (AC) by applying it to the standardly discussed structures that create problems for counterfactual accounts: switching, preemption, overdetermination, and short-circuiting. For reasons of space, I have to focus on applying (AC) to these types of structures and cannot discuss in detail the problems they generate for counterfactual accounts. Yet, all of these problems are well-documented in the literature, most of all, in a recent exchange by Hitchcock and Hall (cf. Hall 2007; Hitchcock 2009).

As Hall (2007, 117-118) shows, structural equations accounts that draw on explicitly nonforetracking counterfactuals identify the firing of F as actual cause of the firing of E in the switching process of figure 3. However, in light of the fact that switch F de facto makes no difference whatsoever to whether neuron E fires, this result conflicts with causal intuitions. (AC), in turn, correctly captures those intuitions. It yields that the firing of F does not count as an actual cause of the firing of E, because the corresponding factor F is not part of a minimal theory of E (over any suitable factor set) and, thus, is no type-level cause of E. By contrast, the firings of A, B, and C in diagram 3 all come out as actual causes of the firing of E. As exhibited in (12), the factors these tokens instantiate are all type-level causes of E and they are connected to E by active causal routes relative to all suitable expansions of $\mathcal{F}_3$ (for, in light of the above completeness assumption, $\mathcal{F}_3$ contains all type-level causes of E). Hence, (AC) correctly mirrors the relations of actual causation that are exhibited by the switching process of figure 3.

Next, let us consider the case of early preemption shown in figure 4a: the firing of neuron C triggers E to fire (via D) and, at the same time, suppresses the stimulation of E by A (via B).[14] In this situation, the actual causes of E's firing are the firings of C and of D. Nonetheless, had C not fired, E would have fired anyway because it then would have been stimulated by A via B. If we model the underlying type-level structure relative to the set $\mathcal{F}_4 = \{A, B, C, D, E\}$ and—as in the previous section—assume orientability of determinis-

[14] Inhibitory signals are represented by '$\multimap$'. They always override stimulatory signals.

tic dependencies, we get the following complex minimal theory:

$$(A\overline{C} \Rightarrow B) \wedge (C \Rightarrow D) \wedge (D \vee B \Rightarrow E) \wedge (A \vee C \Rightarrow E)_i \tag{13}$$

(AC) entails that the instances of C and D in diagram 4a are actual causes of the instance of E. By contrast, the instance of A does not come out an actual cause of E. Although A is part of a minimal theory of E and, thus, a type-level cause of E, A is not connected to E by an active route relative to $\mathcal{F}_4$ in 4a, for A is not co-instantiated with all other members of the minimally sufficient condition $A\overline{C}$ of B. Preempted causes—as the firing of A—are also intuitively not identified as actual causes.

This example demonstrates why (AC) requires tokens to *permanently* connect corresponding factors by active routes in order to be causally related. If the type-level structure instantiated by the process in diagram 4a is modeled relative to the set $\mathcal{F}_4^* = \{A, C, E\}$, $A \vee C \Rightarrow E$ turns out to be the direct minimal theory of E. As a consequence, relative to $\mathcal{F}_4^*$ both A and C are connected to E by active routes. The set $\mathcal{F}_4^*$ is not diverse enough to model the fact that the signal from neuron A to E can be interrupted. Suitably expanding $\mathcal{F}_4^*$ to $\mathcal{F}_4$ yields the richer minimal theory (13) which adequately models the interruptability of that signal and reveals that the firing of A is preempted in the process of diagram 4a.

Cases of late preemption are handled analogously. Since Lewis (1986, 203-204), canonical examples of late preemption have the form of scenario 4b, where neuron E is stimulated by G via C. E suppresses D, which would have triggered E, had E not already received a stimulus along the other path. Modeling the underlying type-level structure relative to the set $\mathcal{F}_5 = \{A, B, C, D, E, G\}$ yields the following minimal theory:

$$\begin{aligned}(G \Rightarrow C) \wedge (A \Rightarrow B) \wedge (B\overline{E} \Rightarrow D) \wedge (C \vee D \Rightarrow E) \wedge \\ (G \vee B \Rightarrow E)_i \wedge (G \vee A \Rightarrow E)_i\end{aligned} \tag{14}$$

A, B, C, and G are all contained in a minimal theory of E. Yet, while G and C additionally are located on an active route to E in diagram 4b, A and B are not. (AC) hence identifies only the firings of G and C as actual causes of the firing of E. Again, this result accords with the usual causal intuitions.[15]

A caveat is required at this point. In recent years, there has been some variance in the literature as to what exactly the difference between early and late preemption amounts to. According to the understanding which hearkens back to Lewis (1986) and which underlies diagrams 4a and 4b, the difference is that "in cases of early preemption, the backup process is cut off before the effect occurs, whereas in cases of late preemption, the process is cut off by the effect itself" (Hitchcock 2007, 526). By contrast, e.g. Hall and Paul (2003, 111-112) hold that the characteristic feature of early preemption is that a process is interrupted by another process, whereas in cases of late preemption no interruption takes place, rather, the preempted process just does not run to completion. Whatever the merits of these two accounts may be, it is clear that in order to adequately reproduce cases of preemption within difference-making theories of causation—be they of the counterfactual or the regularity theoretic type—relevant causal models must contain a factor (or a random variable) that takes on different values depending on whether a corresponding cause is preempted or not (cf. Halpern and Pearl 2005, 862). In a case where, say, two neurons A and C fire at the same time, but the signal of A reaches and triggers neuron E before the signal of C such that the

[15] Keep in mind that (14) is not a propositional expression but a shorthand for a first-order expression that, among other things, imposes spatiotemporal constraints on the instances of the involved factors. In this particular case, these constraints must be taken to imply that $B\overline{E}$ and E are not proximately instantiated (which would be impossible), when neuron E is triggered via D.

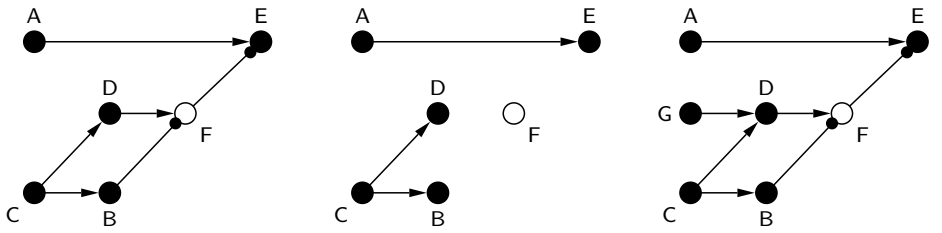


Fig. 5

signal of C does not run to completion, such a *marking* factor (cf. Hitchcock 2004, 416) may e.g. model whether E has already fired or not when the signal from C arrives (cf. Strevens 2007, 103). That is, to reproduce cases of preemption, both structural equations accounts and (AC) require that relevant factor sets are suitably expandable to include at least one marking factor on the preempted causal path.

Diagram 4c features another standard test case for counterfactual accounts: symmetric overdetermination. In this process, neurons C and A trigger E simultaneously, such that each stimulus would have itself been sufficient for E to fire. Intuitively, we want to say that both overdetermining causes count as actual causes of E's firing. (AC) yields this result in a maximally simple manner. Relative to the set $\mathcal{F}_6 = \{A, C, E\}$ the underlying type-level structure is given by the following minimal theory:

$$A \vee C \Rightarrow E \tag{15}$$

That is, both A and C are type-level causes of E and, in diagram 4c, are connected to E by an active causal route each. Thus, the instances of A and C are both identified as actual causes of E by (AC).

Let us now turn to what Hall (2007, 120) has dubbed *short circuits*. An example is given in figure 5a. In this process, neuron C triggers F through D and, at the same time, suppresses F by way of stimulating B. Moreover, F is connected to E through an inhibitory edge, meaning that if F were to fire, E would be suppressed. While structural equations accounts tend to identify the firing of C as an actual cause of E's firing, intuitively the firing of C makes no difference to E at all, because C's stimulatory influence on E's potential suppressor F is canceled by C's own inhibitory signal via B. To see whether (AC) yields the same result, we again have to first identify the relevant minimal theory. Relative to the factor set $\mathcal{F}_7 = \{A, B, C, D, E, F\}$, with factors once more representing the firings of the corresponding neurons, diagram 5a seems to suggest that $A\overline{F}$ is both minimally sufficient and necessary for E. However, on closer inspection, it turns out that in the type-level structure underlying diagram 5a, $A\overline{F}$ has a proper part that is sufficient and necessary for E, *viz.* A, for the other part of $A\overline{F}$, i.e. $\overline{F}$, holds trivially. Under no circumstances could neuron F ever fire, because C and $\overline{C}$ are each minimally sufficient for $\overline{F}$. That is, the tautologous disjunction $C \vee \overline{C}$ determines $\overline{F}$. Therefore, F poses no potential threat to E whatsoever. As neuron F does not possibly make a difference to E, the inhibitory edge between F and E in 5a is ungrounded. Moreover, since C is necessary and sufficient for B and D, it follows that B is instantiated if and only if D is. As a consequence, the type-level dependencies among B, D, and F are

very ambiguous. As a matter of fact, these factors might not be causally connected at all, for the neuron diagram in figure 5a is empirically equivalent to diagram 5b.[16]

Whatever the dependencies among C, B, D, F may be, it is clear that the minimal theory regulating the behavior of E in the type-level structure underlying diagrams 5a and 5b is simply this:

$$A \Rightarrow E \tag{16}$$

As a result, (AC) only identifies the instance of A as an actual cause of the instance of E in the process depicted in 5a and 5b, respectively, and hence accords with causal intuitions.

Matters change radically if we, instead of this simple short circuit, consider the slightly more complex short circuit depicted in diagram 5c. Contrary to 5a (and 5b), 5c features an additional neuron G that can actually cause F to fire. In the process depicted in 5c, the stimulatory influence of G via D on F is suppressed by C through B. In 5c, neuron F poses a real threat for E, and thus there exist circumstances (e.g. the one depicted in 5c) in which firings of C make a difference to whether E fires. In consequence, A is not itself sufficient for E. The type-level structure underlying the behavior of E in 5c involves all factors in the set $\mathcal{F}_8 = \{A, B, C, D, E, F, G\}$. The pertinent minimal theory for diagram 5c is this:

$$\begin{gathered}(C \Rightarrow B) \wedge (C \vee G \Rightarrow D) \wedge (B \vee \overline{D} \Rightarrow \overline{F}) \wedge (A\overline{F} \Rightarrow E) \wedge \\ (C \vee \overline{G} \Rightarrow \overline{F})_i \wedge (AB \vee A\overline{D} \Rightarrow E)_i \wedge (AC \vee A\overline{G} \Rightarrow E)_i\end{gathered} \tag{17}$$

C is now connected to E by an active causal route (via B and $\overline{F}$) and C is moreover a type-level cause of E. Hence, according to (AC), E's firing in diagram 5c is determined to be a joint effect of the firings of C and of A. I consider this result to accord with causal intuitions, for contrary to the process depicted in 5a, the firing of C makes a difference to whether E fires or not in 5c.

Furthermore, I take this result to show that actual causation is not an intrinsic relation of two tokens a and e and, if a is not a direct cause of e, intermediary tokens mediating the causal influence of a on e.[17] Whether two tokens are related in terms of actual causation also hinges on the existence of suitable off-route tokens. Expanding the neuron diagrams 5a and 5b by the additional neuron G turns the firings of C and B into actual causes of E's firing, even though G is not located on the route from C and B to E. The same also holds on the type-level. The contrast between (16) and (17) reveals that adding G turns C and B into type-level causes of E, even though G does not mediate between these factors. The non-intrinsicness of both type and actual causation very naturally follows from analyzing these relations in regularity theoretic terms.

The scenarios in figure 5 also demonstrate that the minimal theories representing the type-level structures underlying neuron diagrams can be changed drastically by integrating (or removing) single neurons. How (AC) analyzes a given example is highly sensitive to the actual complexity of that example. This requires particular caution when comparing the causal claims inferred on the basis of (AC) with an intuitive assessment of a corresponding

[16] Readers with sympathies for interventionism will deny the equivalence of diagrams 5a and 5b by arguing that 5a and 5b do not have the same implications on how E behaves under possible interventions on D or B that are independent of C. According to diagrams 5a and 5b, however, D and B can only be stimulated by C. Hence, there are no possibilities to intervene on D and B independently of C. As will be shown below, as soon as 5a and 5b are suitably expanded by further neurons that can stimulate D or B independently of C the equivalence of 5a and 5b breaks down.

[17] Hall (2004) takes an example analogous to the one in figure 5 to show that there exists at least one concept of causation, *viz. dependence*, that does not amount to an intrinsic relation. Menzies (2002) also significantly weakens his intrinsicness thesis (cf. Menzies 1996) in light of an example of this type.

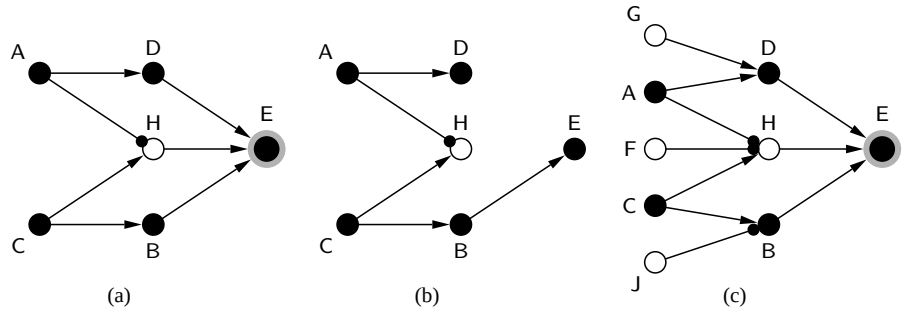


Fig. 6

neuron diagram. The latter must be intuitively assessed without implicitly assuming ways to manipulate certain neurons that are not represented in that diagram.

I conclude this collection of exemplary applications of (AC) with an example that Hall (2004, 263) takes to show that accounts of actual causation in terms material regularities ultimately fail. In diagram 6a, E is a so-called *stubborn* neuron (symbolized by the grey shading) that only fires if it receives at least two stimulatory signals at the same time. In the process depicted in 6a, E is stimulated by A via D and by C via B. A not only stimulates E, but also suppresses H which would have triggered E as well. Diagram 6a suggests that we should count both the firing of A and the firing of C as actual causes of the firing of E. However, if we model the underlying type-level structure relative to $\mathcal{F}_9 = \{A, B, C, D, E, H\}$, it turns out that A is not part of a minimally sufficient condition of E, for E is instantiated if and only if C (and B) is—irrespectively of whether A is also instantiated or not. Under no circumstances could A ever make a difference to E. The type-level structure instantiated by 6a is expressed by the following minimal theory:

$$(A \Rightarrow D) \wedge (C\overline{A} \Rightarrow H) \wedge (C \Rightarrow B) \wedge (B \Rightarrow E) \wedge (C \Rightarrow E)_i \tag{18}$$

Against the background of (18), (AC) of course only identifies the instances of C and B as actual causes of the instance of E in 6a. That is, contrary to what diagram 6a suggests, the firing of A does not come out as cause of the firing of E. Hall (2004) takes this to constitute an insurmountable problem for pure regularity accounts of actual causation.

However, note that expressing the type-level structure instantiated by diagram 6a in terms of the minimal theory (18), first and foremost, reveals that A, D, and H make no difference to E in addition to C and B. Diagram 6a is empirically equivalent to diagram 6b, in which E is not represented as a stubborn neuron and which lacks edges from D to E and from H to E. In view of the fact that causal structures—both on the type and on the token-level—do not feature redundancies, the neuron process in question here should be reproduced in terms of diagram 6b rather than 6a. Obviously, in light of 6b, which does not contain redundant elements, it turns out to be a virtue of (AC) that it does not identify the firing of A as an actual cause of the firing E. In fact, the firing of A is no cause of the firing of E because the former makes no difference whatsoever to the latter. Rather than giving rise to a problem for regularity accounts, the fact that A is not part of a minimally sufficient condition of E reveals that 6a features redundant elements and, hence, does not adequately represent a causal process. Not any graph construed by connecting nodes by stimulatory or inhibitory edges results in a neuron diagram that can be seen to adequately reproduce a causal process.

It will be objected that the stubbornness of E and the capacity of D and H to stimulate E can in fact be tested by suitably intervening on H, D, and B. For instance, if we intervene to suppress H without at the same time stimulating D in a situation where A does not fire, we can test whether the firing of C suffices to trigger E or not, i.e. whether E in fact is stubborn. Similarly, if we can intervene to stimulate D and to suppress B without at the same time suppressing H in a situation where C fires and A does not, we can test whether the firings of D and H indeed make a difference to the behavior of E. Plainly, provided that such interventions are possible the stubbornness of E and the stimulatory impact of D and H on E are easily testable. Diagram 6a, however, does not feature any additional inhibitory and stimulatory neurons that would be required for such intervention tests. Moreover, Hall's argument as to A's failure to be part of a minimally sufficient condition of E essentially hinges on the impossibility to perform these additional interventions. If there indeed exist ways to hold H and B fixed and to stimulate D that are not represented in diagram 6a, relationships of minimal sufficiency change to the effect that A will be part of a minimally sufficient condition of E after all. To see this, consider diagram 6c which results from 6a by integrating additional inhibitory neurons for H and B and an additional stimulatory neuron for D. The minimal theory exhibiting the type-level structure underlying 6c relative to $\mathcal{F}_{10} = \{A, B, C, D, E, F, J, H\}$ is this:

$$(C\overline{J} \Rightarrow B) \wedge (A \vee G \Rightarrow D) \wedge (C\overline{AF} \Rightarrow H) \wedge (BH \vee DH \vee DB \Rightarrow E) \wedge (C\overline{AFJ} \vee GC\overline{AF} \vee AC\overline{J} \vee GC\overline{J} \Rightarrow E)_i \tag{19}$$

In diagram 6c, exactly the same neurons fire as in diagram 6a. However, by integrating the additional neurons required for the intervention tests described above, material regularities change in such way that A is now part of a minimal theory of E and, thus, a type-level cause of E. Moreover, in 6c, A is located on an active route to E. That is, (AC) now rules that the firing of A is an actual cause of the firing of E.

In sum, either diagram 6a is complete or it is not. If it is complete, the firing of A under no possible circumstances makes any difference whatsoever to the behavior of E over and above the firing of C (and B). In that case, diagram 6a is equivalent to 6b. Correspondingly, the firing of A is not determined to be an actual cause of the firing of E by (AC). By contrast, if diagram 6a represents a mere substructure of, say, diagram 6c, there exist circumstances (e.g. the ones represented in 6c) under which the firing of A makes a difference to the behavior of E. In that case, diagram 6a is not equivalent to 6b. Then, A is part of a minimal theory of E and moreover located on an active route to E. Consequently, the firing of A is identified as an actual cause of the firing of E by (AC). In my view, (AC) entails the intuitively adequate relationships of actual causation both if 6a is complete and if it is not.

4 A Relativization to Typicality

An example that is due to Hiddleston (2005) has recently lead to intensified efforts to relativize the notion of actual causation to a context-sensitive standard of normality or typicality (cf. Hitchcock 2007; Hall 2007; Halpern 2008; Halpern and Hitchcock 2010).[18] Consider diagram 7 where E receives an inhibitory signal from C and no stimulatory signal from A.

[18] Instead of typicality, Handfield et al. (2008) relativize actual causation to a context-sensitive condition of salience.

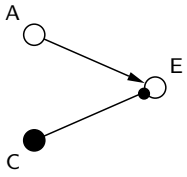


Fig. 7

As a result E does not fire. This process instantiates a type-level structure which, if modeled relative to $\mathcal{F}_{11} = \{A, C, E\}$, entails the regularities expressed in the following minimal theory:

$$\overline{A} \vee C \Rightarrow \overline{E} \tag{20}$$

In diagram 7, both $\overline{A}$ and C are located on an active causal route to $\overline{E}$. Accordingly, both the non-firing of A and the firing of C are identified as actual causes of the non-firing of E by (AC). Structural equations accounts imply the same causal dependencies. However, the intuitive adequacy of this result seems doubtful. In the situation depicted in diagram 7, A does not fire. Thus, the inhibitory signal E receives from C appears to be completely irrelevant. That is, causal intuitions tend to identify the non-firing of A as the only actual cause of the non-firing of E.[19]

Diagram 7 is structurally isomorphic to diagram 4c. While the latter exhibits a case of an overdetermined occurrence, the former depicts an overdetermined absence. Unsurprisingly, the corresponding minimal theories (15) and (20) are isomorphic as well. However, while in the case of 4c causal intuitions clearly identify both overdetermining tokens as actual causes, intuitions tend towards a different assessment in case of 7. Hitchcock, Hall, and Halpern take this to show that actual causation does not only depend on the counterfactual dependencies that are implied by corresponding causal processes and that are encoded in structural equations. Assessments of actual causation additionally depend on "a theory of 'normality' or 'typicality"' (Halpern 2008, 204).

Plainly, relativizing actual causation to typicality standards renders actual causation dependent on pragmatic features of the concrete context in which a causal process is modeled. Yet, another widespread causal intuition has it that whether or not two tokens are causally related is an entirely objective matter which in no way hinges on contingencies of modeling contexts. Therefore, rather than taking the conflict between the intuitive assessments of diagrams 4c and 7 to count against the context-independence of actual causation, I would prefer to take this conflict to reveal a confusion in our causal intuitions. Whoever hesitates to acknowledge that the instance of C in diagram 7 is an actual cause of $\overline{E}$ in fact has a pragmatic and context-dependent causal notion in mind, most likely *causal explanation*. In the end, however, whether or not Hiddleston-type examples are interpreted to show that the notion of actual causation must be contextualized and approximated to the notion of causal explanation is a terminological issue over which I do not want to argue. It is a fact that many authors opt for contextualization. Therefore, I conclude this paper by briefly indicating how the contextualization techniques adopted in the structural equations framework can also be used to define a regularity theoretic notion of actual causation that is relativized to a standard of typicality.

Factors that are connected by a deterministic causal structure can be instantiated in certain configurations and not in others. For instance, according to the type-level structure underlying diagram 7 exactly the configurations listed in table 1a are empirically possible.

[19] These intuitions are further strengthened by changing the interpretation of the occurrences in 7, as e.g. done in a scenario Hitchcock (2007, 523) calls *Bogus Prevention*.

That is, A and C can be instantiated while E is not (configuration c_1), or C can be instantiated while A and E are not (c_2), or all three factors can be absent (c_3), or A and E can be instantiated while C is not (c_4). All other logically possible configurations of the factors in $\mathcal{F}_{11} = \{A, C, E\}$ are determined to be empirically impossible by the type-level structure behind diagram 7. Minimal theories simply express the configurations listed in table 1a in a standardized syntactic form. Accordingly, the minimal theory (20) is true if, and only if, the factors in $\mathcal{F}_{11}$ take one of the value configurations listed in table 1a.

In ordinary contexts of causal modeling, not all possible value configurations for an analyzed structure are equally typical. Hence, relativizing (AC) to contextually induced typicality rankings, first and foremost, presupposes that possible value configurations are ordered according to the typicality ranking that is relevant for a given modeling context. In the structural equations framework, it is customary to assign the lowest rank to the most typical configuration and to increase the rank with decreasing typicality. A token a can then be said to be a contextualized actual cause of another token e iff a and e satisfy (AC) and a additionally makes a difference to e relative to the configurations with equal or lower typicality rank than the actual configuration (cf. Halpern 2008).

In order to make this idea somewhat more precise, consider table 1b which exhibits one conceivable ranking of the configurations listed in table 1a. The scenario depicted in diagram 7 is of type c_2 and, according to the ranking of table 1b, is of typicality rank 2. To determine whether the firing of C in diagram 7 makes a difference to the non-firing of E relative to all configurations with equal or lower rank than c_2, we first eliminate the one possible configuration with higher rank, i.e. c_1, and second, check whether the corresponding factor C is still part of a minimally necessary condition of $\overline{E}$ relative to this truncated list of configurations. Table 1c constitutes such a truncation of 1b. As can easily be seen, relative to the configurations in table 1c, C is not part of a minimally necessary condition of $\overline{E}$ any more. In configurations c_2 and c_3 of 1c the value of C changes while both $\overline{A}$ and $\overline{E}$ remain unchanged. Hence, relative to the configurations with a maximal rank of 2, C makes no difference to $\overline{E}$. The minimal theory expressing the configurations in 1c is this:

$$\overline{A} \Rightarrow \overline{E} \tag{21}$$

We might call (21) a *contextually weighted* minimal theory for diagram 7. It reproduces the relations of minimal sufficiency and necessity holding among the factors in $\mathcal{F}_{11}$ relative to the set of configurations with equal or lower typicality rank than the configuration in the actual situation, i.e. in diagram 7. Against this background, a contextualized notion of actual causation can be more precisely defined as follows: a token a is a *contextualized actual cause* of a token e iff a and e satisfy (AC) and, relative to a factor set $\mathcal{F}_i$ that is used in a given modeling context and that contains factors A and E such that A is instantiated by a and E by e, A is part of a contextually weighted minimal theory Φ_i of E over $\mathcal{F}_i$. Thus, the firing

#	A	C	E
c_1	1	1	0
c_2	0	1	0
c_3	0	0	0
c_4	1	0	1

(a)

#	A	C	E	rank
c_1	1	1	0	3
c_2	0	1	0	2
c_3	0	0	0	1
c_4	1	0	1	2

(b)

#	A	C	E	rank
c_2	0	1	0	2
c_3	0	0	0	1
c_4	1	0	1	2

(c)

Table 1

of C in diagram 7 is no contextualized actual cause of the non-firing of E because C is not contained in the contextually weighted minimal theory (21).

Of course, this is only a rough sketch of a regularity theoretic notion of actual causation that is relativized to typicality standards. Nonetheless, it should suffice to substantiate that, if desired, a regularity theory can be relativized to such standards along analogous lines as structural equations accounts.

5 Conclusion

This paper has shown that in order to account for token-level processes contained in the standard set of test cases no recourse to nonforetracking counterfactuals nor even to non-actual possible worlds is required. Preemption, overdetermination, switching, and short-circuiting—all of which cause problems for some counterfactual analyses or other—can be accounted for on the basis of rigorously minimized material regularities that are permanent across extensions of causally modeled factor sets. As anticipated in the introduction, I do not claim that (AC) is beyond doubt in all conceivable cases. For instance, I did not discuss cases of trumping (cf. Schaffer 2000) or of preemptive prevention (cf. Collins 2000). There are different intuitions as to how to assess these structures. Hitchcock (2007, 512) treats trumping as a species of overdetermination and preemptive prevention as a species of early preemption (cf. also McDermott 2002; Halpern and Pearl 2005). If treated as such, they do neither constitute a problem for structural equations accounts nor for (AC). However, Schaffer (2000) and Collins (2000) hold that trumping and preemptive prevention are not reducible to overdetermination and preemption. In that case, they might well turn out to give rise to problems both for modern counterfactual accounts and for (AC). Overall, I only want to claim that the latter performs at least as well as the former. Moreover, contrary to theories employing structural equations, (AC) achieves its goal by implementing uncontroversial and straightforward conceptual and technical resources only. The ease with which structures of actual causation that create problems for the structural equations framework can be properly reproduced in a regularity theoretic framework should be reason enough to take regularity theories more seriously than they are currently taken.

Acknowledgements I thank Luke Glynn, Wolfgang Spohn, and two anonymous referees of this journal for very helpful comments on earlier drafts. Moreover, I have profited a lot from discussions with audiences at two workshops held at the University of Konstanz in 2009/10. Finally, I am indebted to the Deutsche Forschungsgemeinschaft (DFG) for generous support of this work (project CAUSAPROBA).

References

Armstrong, D. M. (1983). *What is a Law of Nature?* Cambridge: Cambridge University Press.

Baldwin, R. A. and E. Neufeld (2004). The structural model interpretation of the NESS test. In *Advances in Artificial Intelligence*, Volume 3060, pp. 297–307.

Baumgartner, M. (2008). Regularity theories reassessed. *Philosophia 36*, 327–354.

Baumgartner, M. (2009). Uncovering deterministic causal structures: a Boolean approach. *Synthese 170*, 71–96.

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

Collins, J. (2000). Preemptive prevention. *Journal of Philosophy 97*, 223–234.

Collins, J., N. Hall, and L. Paul (Eds.) (2004). *Causation and Counterfactuals*, Cambridge. MIT Press.

Fodor, J. (1997). Special sciences: still autonomous after all these years. *Noûs 31*, 149–163.

Glymour, C., D. Danks, B. Glymour, F. Eberhardt, J. Ramsey, R. Scheines, P. Spirtes, C. Teng, and J. Zhang (2010). Actual causation: a stone soup essay. *Synthese 175*, 169–192.

Graßhoff, G. and M. May (2001). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.
Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, and L. Paul (Eds.), *Counterfactuals and Causation*, pp. 225–276. Cambridge: MIT Press.
Hall, N. (2007). Structural equations and causation. *Philosophical Studies 132*, 109–136.
Hall, N. and L. A. Paul (2003). Causation and preemption. In P. Clark and K. Hawley (Eds.), *Philosophy of Science Today*, pp. 100–130. Oxford: Oxford University Press.
Halpern, J. Y. (2008). Defaults and normality in causal structures. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning*, pp. 198–208.
Halpern, J. Y. and C. Hitchcock (2010). Actual causation and the art of modelling. In R. Dechter, H. Geffner, and J. Y. Halpern (Eds.), *Heuristics, Probability, and Causality*, pp. 383–406. London: College Publications.
Halpern, J. Y. and J. Pearl (2005). Causes and explanations: a structural-model approach. Part I: Causes. *British Journal for the Philosophy of Science 56*, 843–887.
Handfield, T., C. R. Twardy, K. B. Korb, and G. Oppy (2008). The metaphysics of causal models: where's the biff? *Erkenntnis 68*(2), 149–68.
Hausman, D. (1998). *Causal Asymmetries*. Cambridge: Cambridge University Press.
Hiddleston, E. (2005). Causal powers. *British Journal for the Philosophy of Science 56*, 27–59.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy 98*, 273–299.
Hitchcock, C. (2004). Do all and only causes raise the probabilities of effects? In J. Collins, N. Hall, and L. Paul (Eds.), *Causation and Counterfactuals*, pp. 403–417. MIT Press.
Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review 116*, 495–532.
Hitchcock, C. (2009). Structural equations and causation: six counterexamples. *Philosophical Studies 144*, 391–401.
Hitchcock, C. (2010). Probabilistic causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2010 ed.).
Kim, J. (1971). Causes and events: Mackie on causation. *Journal of Philosophy 68*, 426–441.
Lewis, D. (1973). Causation. *Journal of Philosophy 70*, 556–567.
Lewis, D. (1986). Postscript to 'causation'. In *Philosophical Papers*, Volume 2, pp. 172–213. Oxford: Oxford University Press.
Lewis, D. (1999). New work for a theory of universals. In *Papers in Metaphysics and Epistemology*, pp. 8–55. Cambridge: Cambridge University Press.
Mackie, J. L. (1974). *The Cement of the Universe. A Study of Causation*. Oxford: Clarendon Press.
Mackie, J. L. (1993 (1965)). Causes and conditions. In E. Sosa and M. Tooley (Eds.), *Causation*, pp. 33–55. Oxford: Oxford University Press.
Maudlin, T. (2004). Causation, counterfactuals, and the third factor. In J. Collins, N. Hall, and L. Paul (Eds.), *Causation and Counterfactuals*, pp. 419–443. MIT Press.
May, M. (1999). *Kausales Schliessen. Eine Untersuchung über kausale Erklärungen und Theorienbildung.* Ph. D. thesis, Universität Hamburg, Hamburg.
McDermott, M. (2002). Causation: influence versus sufficiency. *The Journal of Philosophy 99*, 84–101.
Menzies, P. (1996). Probabilistic causation and the pre-emption problem. *Mind 105*, 85–117.
Menzies, P. (2002). Is causation a genuine relation? In G. Rodriguez-Pereya and H. Lillehammer (Eds.), *Real Metaphysics. Essays in Honor of D. H. Mellor*, pp. 120–136. London: Routledge.
Pearl, J. (2000). *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Psillos, S. (2009). Causation and regularity. In H. Beebee, P. Menzies, and C. Hitchcock (Eds.), *Oxford Handbook of Causation*, pp. 131–157. Oxford: Oxford University Press.
Quine, W. v. O. (1959). On cores and prime implicants of truth functions. *The American Mathematical Monthly 66*, 755–760.
Schaffer, J. (2000). Causation by disconnection. *Philosophy of Science 67*, 285–300.
Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2 ed.). Cambridge: MIT Press.
Strevens, M. (2007). Mackie remixed. In J. K. Campbell, M. O'Rourke, and H. S. Silverstein (Eds.), *Causation and Explanation*, Volume 4 of *Topics in Contemporary Philosophy*, pp. 93–118. Cambridge: MIT Press.
Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.
Wright, R. W. (1985). Causation in tort law. *California Law Review 73*, 1735–1828.