Proceedings of the Sixteenth International Conference on
Principles of Knowledge Representation and Reasoning (KR 2018)

On Laws and Counterfactuals in Causal Reasoning

Alexander Bochman
Computer Science Department,
Holon Institute of Technology, Israel
bochmana@hit.ac.il

Abstract

We explore the relationships between causal rules and counterfactuals, as well as their relative representation capabilities, in the logical framework of the causal calculus. It will be shown that, though counterfactuals are readily definable on the basis of causal rules, the reverse reduction is achievable only up to a certain logical threshold (basic equivalence). As a result, we will argue that counterfactuals cannot distinguish causal theories that justify different claims of actual causation, which could be seen as the main source of the problem of 'structural equivalents' in counterfactual approaches to causation. This will lead us to a general conclusion about the primary role of causal rules in representing causation.

Introduction

Causation plays a crucial role in our view of the world, from commonsense reasoning to natural and social sciences, and up to jurisprudence, linguistic semantics and AI formalisms. Accordingly, it should occupy an appropriate place in general Knowledge Representation. One of the essential prerequisites of such a representation, however, is the choice of basic informational units and adequate formalisms for encoding causal knowledge. These choices have turned out to be both non-trivial and controversial.

Traditionally, the triple of notions *causation*, *counterfactuals* and *laws* has been at the heart of the philosophy of science, and the relations between them have been the focus of much discussions and controversy. An undeniable basis of these discussions, however, has always been the fact that these three notions are tightly correlated, or, using David Lewis phrase, they are "rigidly fastened to one another, swaying together rather than separately". Such a correlation obviously required explanation, and many approaches have been suggested that tried to define each of these concepts in terms of the rest. Hume's informal analysis of causation in terms of regularities (laws) and/or counterfactuals, and David Lewis' counterfactual analysis of causation (Lewis 1973) have been the most prominent approaches of this reductive kind. More recently, the framework of structural equation models (Pearl 2000; Spirtes, Glymour, and Scheines 2000) has provided a rigorous basis for reasoning with these concepts, but it has not ended the controversy.

According to (Pearl 2000), the basic building blocks of the structural account of causation are structural equations, which are functions that represent lawlike mechanisms. These equations can be naturally viewed as formal counterparts of (causal) *laws*, since they describe generic (type-level) relations among variables that are applicable to every hypothetical scenario. That is why they are capable of determining corresponding counterfactuals (via the notions of intervention and sub-model). Speaking more generally, structural equations provide information necessary for supporting all kinds of causal claims and, in particular, the claims of actual causation[1].

Though in Pearl's approach structural equations were taken as primitive, an influential camp of philosophers and researchers has continued David Lewis' legacy in arguing that counterfactuals, in one way or other, should still enjoy a principal status in causal reasoning. In its simplest form, the argument has been that the structural equations themselves just represent certain privileged counterfactuals[2]. In a more elaborate approach of (Woodward 2003), the main objective was to provide a manipulative (counterfactual) account of the causal notions that Pearl has taken as primitive. According to Woodward, facts about patterns of counterfactual dependence are more basic than facts about what causes what, and the essence of the manipulability account can be put in a slogan "No causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference".

It seems that this line of thought has also influenced the bulk of recent counterfactual approaches to actual causation in the framework of structural equations, including the HP definitions of Halpern and Pearl (Halpern and Pearl 2001; 2005; Halpern 2016a).

(Bochman and Lifschitz 2015) has suggested a logical representation of Pearl's causal models in the causal calculus (McCain and Turner 1997; Lifschitz 1997). In this representation, structural equations were 'translated' to causal rules,

[1] The initial definition of actual causation in (Pearl 2000, Chapter 10) used the notions of sustenance and causal beam that were directly defined in terms of structural equations.

[2] (Hitchcock 2007): "counterfactuals are represented using equations among the variables, where each equation asserts several counterfactuals: one for each assignment of values to the variables that makes the equations true."

and it has been shown, in particular, how interventions and submodels can be described in this framework. In addition, it has been shown in (Bochman 2018) that actual causation can be directly defined in the causal calculus, without intermediate help from counterfactuals.

In this study we are going to recast the philosophical 'trilemma' of laws, counterfactuals and causation in the logical framework of the causal calculus, which will allow us to derive some precise conclusions about their relationships.

As a first step, we will formally define counterfactuals in the causal calculus. It will be shown, in particular, that checking the validity of counterfactuals on this definition amounts to verifying classical entailment in certain completions of the source causal theory.

The suggested definition of counterfactuals will allow us to pose a precise question to what extent counterfactuals can capture the 'causal content' of the source causal theory. In this respect, it will be shown that there exists a logical threshold for such a reduction, namely logical equivalence with respect to the basic causal inference (see below). In other words, basically equivalent causal theories are indistinguishable by counterfactual tools, since they support the same counterfactuals.

As a next step, we will connect the above results with the observations made in (Bochman 2018), according to which basically equivalent causal theories can support different claims of actual causation. This fact could even be seen as the source of the well-known problem of 'structural equivalents' in counterfactual theories of causation.

Finally, we will connect the lessons that could be learned from the above results with a general approach to the trilemma of causation, counterfactuals and laws suggested in (Maudlin 2004).

General Causation in the Causal Calculus

Originally, the causal calculus has been introduced in (McCain and Turner 1997) as a nonmonotonic formalism for reasoning about action and change in AI (see (Giunchiglia et al. 2004)). A logical basis of the causal calculus was described in (Bochman 2003), while (Bochman 2004) studied its possible uses as a general-purpose nonmonotonic formalism.

In this study, we will use the causal calculus as a general logical formalism of causal reasoning. As such, it shares a common starting point with Pearl's approach to causality in that our knowledge can be stored in terms of cause-effect relationships. In the causal calculus, the latter are represented directly by causal rules of the form $A \Rightarrow B$ (*"A causes B"*), where A and B are classical propositions. Structural equation models are representable using such rules, so the approach can be viewed as a logical generalization of the latter.

Causal rules represent general (type-level) causal claims, so they correspond to such notions as nomic or causal sufficiency, causal laws and lawlike regularities. Just as the latter, causal rules are inherently *modal* notions.

As in Pearl's approach, causal rules will be viewed as representing causal *mechanisms*, though our representation will be based on a more fine-grained understanding of mechanisms than what is usually assumed in structural equation models.

Our basic language will be an ordinary propositional language with the classical connectives and constants $\{\wedge, \vee, \neg, \rightarrow, \mathbf{t}, \mathbf{f}\}$. $\vDash$ will stand for the classical entailment, while Th will denote the classical provability operator. We will often identify a propositional interpretation ('world') with the set of propositional formulas that hold in it.

In what follows, by a *causal theory* we will mean a set of causal rules. A causal theory will be called *determinate* if it contains only rules of the form $A \Rightarrow l$, where l is a literal. Throughout this study, we will restrict our attention to finite determinate causal theories, and some of the key results below will depend on this restriction.

Nonmonotonic Semantics

A distinctive feature of causal reasoning is that situations described by a causal theory are determined not only by the rules that belong to the theory, but also by what does *not* belong to it. Accordingly, this principal semantic function is realized in the causal calculus by assigning a causal theory a particular *nonmonotonic* semantics: the relevant situations should not only be closed with respect to the causal rules of the theory, they should also satisfy Leibnitz's Principle of Sufficient Reason: nothing happens without a sufficient reason, why it should be so. Formally, a nonmonotonic semantics of a causal theory can be defined as follows.

For a causal theory Δ and a set u of propositions, let $\Delta(u)$ denote the set of propositions that are caused by u in Δ:

$$\Delta(u) = \{B \mid A \Rightarrow B \in \Delta, \text{ for some } A \in u\}$$

Definition 1.
- A set u of propositions is an *exact model* of a causal theory Δ if it is consistent, and $u = \mathrm{Th}(\Delta(u))$.
- A *general nonmonotonic semantics* of a causal theory is the set of all its exact models.
- A *causal nonmonotonic semantics* of a causal theory is the set of its exact models that are worlds (complete deductively closed sets).

An exact model describes an information state that is closed with respect to the causal rules, but in which also every proposition is caused, or *explained*, by other propositions that hold in this state.

The causal nonmonotonic semantics of causal theories is equivalent to the semantics described in (McCain and Turner 1997) and used in (Giunchiglia et al. 2004).

Regular, Basic and Causal Inference

The causal calculus can be viewed as a two-layered construction. The nonmonotonic semantics defined above form its top level. Its bottom level are a number of logics for causal rules introduced in (Bochman 2003; 2004); they function as *causal logics* of the causal calculus.

The following general notion of production inference is actually a slight modification of the input-output logic from (Makinson and van der Torre 2000).

Definition 2. A *production inference relation* is a binary relation $\Rightarrow$ on the set of classical propositions satisfying the following conditions:

(Strengthening) If $A \vDash B$ and $B \Rightarrow C$, then $A \Rightarrow C$;

(Weakening) If $A \Rightarrow B$ and $B \vDash C$, then $A \Rightarrow C$;

(And) If $A \Rightarrow B$ and $A \Rightarrow C$, then $A \Rightarrow B \wedge C$;

(Truth) $\mathbf{t} \Rightarrow \mathbf{t}$;

(Falsity) $\mathbf{f} \Rightarrow \mathbf{f}$.

A characteristic property of production inference is that the reflexivity postulate $A \Rightarrow A$ does not hold for it.

We extend causal rules to rules having arbitrary sets of propositions as premises by employing compactness: for any set u of propositions, we define

$$u \Rightarrow A \;\equiv\; \bigwedge a \Rightarrow A, \text{ for some finite } a \subseteq u$$

$\mathcal{C}(u)$ denotes the set of propositions caused by u, that is

$$\mathcal{C}(u) = \{A \mid u \Rightarrow A\}$$

As could be expected, the causal operator $\mathcal{C}$ plays much the same role as the usual derivability operator for consequence relations. Note that $\mathcal{C}(u)$ is always a deductively closed set. In addition, it is a monotonic, and even continuous, operator. Still, it is not inclusive, that is, $u \subseteq \mathcal{C}(u)$ does not always hold. Also, it is not idempotent, that is, $\mathcal{C}(\mathcal{C}(u))$ can be distinct from $\mathcal{C}(u)$.

A production inference relation is *regular* if it satisfies the following well-known rule:

(Cut) If $A \Rightarrow B$ and $A \wedge B \Rightarrow C$, then $A \Rightarrow C$.

Cut is one of the basic rules for ordinary consequence relations. In the context of production inference it plays the same role, namely, allows for a reuse of produced propositions as premises in further derivations. It is important to note, in particular, that regular inference relations are already transitive.

Regular inference relations have played an important role in describing actual causation in (Bochman 2018).

Following (Makinson and van der Torre 2000), a production inference relation is called *basic* if it satisfies

(Or) If $A \Rightarrow C$ and $B \Rightarrow C$, then $A \vee B \Rightarrow C$.

For basic production inference, the set of propositions caused by a propositional theory coincides with the set of propositions that are caused by every world containing it:

$$\mathcal{C}(u) = \bigcap\{\mathcal{C}(\alpha) \mid u \subseteq \alpha \;\&\; \alpha \text{ is a world}\}$$

Another important fact about basic production inference is that any causal rule is reducible to a set of *clausal* rules of the form $\bigwedge l_i \Rightarrow \bigvee l_j$, where l_i, l_j are classical literals.

The following characterization of basic equivalence for determinate causal theories is used in proofs of the results below.

A world α will be called *causally consistent* with respect to a causal theory Δ if $\Delta(\alpha)$ is a classically consistent set.

Lemma 1. *Determinate causal theories Δ and Γ are basically equivalent iff they have the same causally consistent worlds, and $\Delta(\alpha) = \Gamma(\alpha)$, for any causally consistent world α.*

Finally, a production inference relation will be called *causal* if it is both basic and regular. Causal inference relations satisfy almost all the usual postulates of classical inference (except Reflexivity and Contraposition).

In what follows, $\Rightarrow_\Delta$ will denote by the 'causal closure' of a causal theory Δ, namely the least causal inference relation that includes Δ, whereas $\Rightarrow^r_\Delta$ and $\Rightarrow^b_\Delta$ will denote, respectively, the least regular and the least basic production relations that include Δ. Each of these inference relations will play an important role in this study.

Completion

The causal nonmonotonic semantics of a determinate causal theory Δ coincides with the classical semantics of the propositional theory obtained from Δ by a syntactic transformation similar to program completion.

The *completion* of a (finite) determinate causal theory Δ is the set $comp(\Delta)$ of all classical formulas of the forms

$$p \leftrightarrow \bigvee\{A \mid A \Rightarrow p \in \Delta\}$$
$$\neg p \leftrightarrow \bigvee\{A \mid A \Rightarrow \neg p \in \Delta\},$$

for any atom p, plus the set $\{\neg A \mid A \Rightarrow \mathbf{f} \in \Delta\}$.

As proved in (McCain and Turner 1997), the completion of a determinate causal theory provides a classical logical description of its nonmonotonic semantics:

Proposition 2. *The causal nonmonotonic semantics of a determinate causal theory coincides with the classical semantics of its completion.*

It should be kept in mind, however, that this completion transformation is not modular with respect to the causal rules of the source theory and, moreover, it changes nonmonotonically with the changes of the latter. Speaking generally, the completion (as well as the nonmonotonic semantics itself) does not fully represent the *logical* content of a causal theory. This distinction between logical and nonmonotonic aspects of a causal theory bears immediate relevance to the distinction between causal and purely mathematical understanding of structural equations in Pearl's theory of causality that we will briefly describe in the next section.

Representing Structural Equations

According to (Pearl 2000, Chapter 7), a causal model is a triple $M = \langle U, V, F\rangle$ where U is a set of *exogenous* variables, V is a finite set of *endogenous* variables, and F is a set of functions such that each $f_i \in F$ is a mapping from $U \cup (V \backslash V_i)$ to V_i.

Symbolically, F is represented as a set of equations

$$v_i = f_i(pa_i, u_i) \quad i = 1, \ldots, n$$

where pa_i is any realization of the unique *minimal* set of variables PA_i in $V\backslash\{V_i\}$ (parents) sufficient for representing f_i, and similarly for $U_i \subseteq U$.

Every instantiation $U = u$ of the exogenous variables determines a "causal world" of the causal model. Such worlds stand in one-to-one correspondence with the solutions to the

above equations in the ordinary mathematical sense. However, structural equations also encode causal information in their very syntax by treating the variable on the left of $=$ as the effect and treating those on the right as causes. This causal reading plays a crucial role in determining the effect of external interventions and evaluation of counterfactuals. Each structural equation is intended to represent a stable and autonomous physical mechanism, which means that it is conceivable to modify (or cancel) one such equation without changing the others. In particular, in order to answer counterfactual queries, we have to consider *submodels* of a given causal model. Given a particular instantiation x of a set of variables X from V, a submodel M_x of M is obtained from M by replacing all functions f_i corresponding to members of set X with the set of constant functions $X = x$.

For binary variables, Pearl's notion of a model can be formulated as follows (cf. (Bochman and Lifschitz 2015)):

Definition 3. Assume that the set of propositional atoms is partitioned into a set of *exogenous* atoms and a finite set of *endogenous* atoms.

- A *Boolean structural equation* is an expression of the form $A = F$, where A is an endogenous atom and F is a propositional formula in which A does not appear.
- A *Boolean causal model* is a set of Boolean structural equations $A = F$, one for each endogenous atom A.

Definition 4. A *solution* (or a *causal world*) of a Boolean causal model M is any propositional interpretation satisfying the equivalences $A \leftrightarrow F$ for all equations $A = F$ in M.

(Bochman and Lifschitz 2015) suggested the following translation of causal models into the causal calculus.

Definition 5. For any Boolean causal model M, Δ_M is the causal theory consisting of the rules

$$F \Rightarrow A \text{ and } \neg F \Rightarrow \neg A$$

for all equations $A = F$ in M and the rules

$$A \Rightarrow A \text{ and } \neg A \Rightarrow \neg A$$

for all exogenous atoms A of M.

The above representation faithfully reflects the source description in structural models by which both truth and falsity assignments to an endogenous atom should be determined by the corresponding function.

Proposition 3 ((Bochman and Lifschitz 2015))**.** *The causal worlds of a Boolean causal model M are identical to the exact worlds of Δ_M.*

Example 1. In the 'firing squad' example from (Pearl 2000, Chapter 7), let U, C, A, B, D stand, respectively, for the following propositions: "Court orders the execution", "Captain gives a signal", "Rifleman A shoots", "Rifleman B shoots", and "Prisoner dies." The story is formalized using the following causal model M, in which U is the only exogenous atom:

$$\{C = U,\ A = C,\ B = C,\ D = A \vee B\}.$$

It has two solutions: in one of them all atoms are true, in the other all atoms are false. This causal model allows us to answer 'static' queries concerning the domain. It corresponds to the following causal theory Δ_M:

$$U \Rightarrow C,\ \neg U \Rightarrow \neg C,\ C \Rightarrow A,\ \neg C \Rightarrow \neg A,$$
$$C \Rightarrow B,\ \neg C \Rightarrow \neg B,\ A \vee B \Rightarrow D,\ \neg(A \vee B) \Rightarrow \neg D,$$
$$U \Rightarrow U,\ \neg U \Rightarrow \neg U.$$

This causal theory has two exact models, identical to the solutions (causal worlds) of M.

As has been noted in (Bochman and Lifschitz 2015), the representation of causal models in the causal calculus produces a particular class of causal theories that is subsumed by the following definition:

Definition 6. A causal theory will be called a *causal Pearl theory* if it is determinate and satisfies the following conditions:

- no atom can appear both in the head and the body of a causal rule;
- two rules $A \Rightarrow p$ and $B \Rightarrow \neg p$ belong to a causal theory only if $A \wedge B$ is classically inconsistent.

This particular class of causal theories will play an important role in our subsequent results.

Counterfactuals in the Causal Calculus

As a starting point of this study, we will provide a formal definition of counterfactuals in the causal calculus.

In the framework of Pearl's causal models, counterfactuals are defined using the notions of intervention and submodel. In the translation of (Bochman and Lifschitz 2015), the latter correspond to sub-theories of a causal theory. We will provide below a somewhat different definition of these notions that will exploit a striking similarity between interventions and belief revision operations. However, due to some well-known difficulties in defining interventions with respect to arbitrary logical formulas, these 'causal revisions', as well as the corresponding counterfactuals, will be restricted to literals in this study.

For a set L of literals, we will denote by $\neg L$ the set of classical literals corresponding to $\{\neg l \mid l \in L\}$.

Definition 7. Given a determinate causal theory Δ and a set L of literals,

- the *contraction* $\Delta - L$ of Δ with respect to L is the determinate causal theory obtained from Δ by removing all rules $A \Rightarrow l$ for $l \in L$.
- The *revision* $\Delta * L$ of Δ is the determinate causal theory obtained from the contraction $\Delta - \neg L$ by adding the rule $\mathbf{t} \Rightarrow l$ for each $l \in L$.

It can be immediately verified that revisions of causal theories exactly correspond to submodels of Boolean causal models (as defined in (Bochman and Lifschitz 2015)).

Example 1, continued. In the firing squad example, let us consider the following action sentence (in the terminology of (Pearl 2000)):

S4. If the captain gave no signal and rifleman A decides to shoot, the prisoner will die and B will not shoot.

To evaluate it, we have to consider the revision $\Delta * \{\neg C, A\}$ of the original causal theory:

$$\mathbf{t} \Rightarrow \neg C,\ \mathbf{t} \Rightarrow A,$$
$$C \Rightarrow B,\ \neg C \Rightarrow \neg B,\ A \vee B \Rightarrow D,\ \neg(A \vee B) \Rightarrow \neg D,$$
$$U \Rightarrow U,\ \neg U \Rightarrow \neg U.$$

Both D and $\neg B$ hold in all causal worlds of the above theory, so S4 is justified.

By a *counterfactual* we will mean an expression of the form $L{>}A$, where L is a finite set of literals, and A a proposition. Traditionally, counterfactuals are defined semantically with respect to worlds. The interventionist definition suggests, however, a powerful and useful generalization of validity for counterfactuals with respect to causal theories.

Definition 8. Counterfactual $L > A$ will be said to *hold in a causal theory* Δ (notation $L{>}_{\Delta}A$), if A holds in all causal worlds of the revision $\Delta * L$.

As in the structural account, acyclic causal theories always determine a unique causal world for any interpretation of the exogenous variables. Accordingly, given a causal world α of a causal theory Δ, let Δ^{α} be the causal theory obtained from Δ by adding rules $\mathbf{t} \Rightarrow l$ for each exogenous literal $l \in \alpha$.

Definition 9 (World-based counterfactuals)**.** Counterfactual $L > B$ will be said to *hold in a causal world* α of a causal theory Δ if it holds in Δ^{α}.

It can be easily verified that the above definition coincides, in effect, with the standard definition of counterfactuals in structural equation models.

Example 1, continued. Following (Pearl 2000), given the actual world $\alpha = \{U, C, A, B, D\}$ of the firing squad example in which the prisoner is dead, let us evaluate the following counterfactual $\neg A > D$:

The prisoner would be dead even if rifleman A had not shot.

By our definition, this world-based counterfactual should be evaluated with respect to Δ^{α}, which can be safely reduced to the following causal theory:

$$\mathbf{t} \Rightarrow U,\ U \Rightarrow C,\ C \Rightarrow A,\ C \Rightarrow B,\ A \vee B \Rightarrow D.$$

The revision of this causal theory with $\neg A$, $\Delta^{\alpha} * \neg A$, is

$$\mathbf{t} \Rightarrow U,\ U \Rightarrow C,\ \mathbf{t} \Rightarrow \neg A,\ C \Rightarrow B,\ A \vee B \Rightarrow D.$$

The latter causal theory has a unique causal world $\{U, C, \neg A, B, D\}$ where D holds, so $\neg A > D$ holds in α.

As can be seen, the above definitions provide feasible tools for evaluating counterfactuals both with respect to specific worlds and causal theories in general. Furthermore, the completion construction for determinate causal theories, mentioned earlier, can be adapted to the case of counterfactuals, which will give us the following key result:

Theorem 4. *A counterfactual $L > B$ holds in a determinate causal theory Δ if and only if*

$$comp(\Delta{-}\neg L) \models \wedge L \rightarrow B.$$

The above result reduces, in effect, checking counterfactual assertions in the causal setting to classical entailment.

Causal Diagrams and Parsimony

In the following sections we are going to explore the expressive capabilities of the counterfactual language. As a first step, we will re-establish an important positive claim of the structural account that counterfactuals are sufficient for determining direct causes and the causal diagram associated with a causal theory.

On the structural account, each endogenous variable is determined by a unique structural equation, which is purported to represent a single underlying causal mechanism. Each variable appearing on the right side of the equation ('parent') is viewed then as a direct cause of this endogenous variable. This understanding presupposes, however, that the relevant equation does not contain redundant variables that do not influence the output. Formally, this restriction can be articulated using the following *test pair condition* (see, e.g., (Glymour et al. 2010)):

Definition 10. For each parent X of a variable Y, the function $Y = f(Parents(Y))$ allows a *test pair* for X with respect to Y, if there exist two causal worlds, α and β, such that (i) for all variables Z in $Parents(Y) \setminus X$, $\alpha(Z) = \beta(Z)$; (ii) $\beta(X) \neq \alpha(X)$; and (iii) $f(\alpha(Parents(Y)) \neq f(\beta(Parents(Y))$.

The above condition requires that a parent variable must actually matter (i.e., make a difference) for its effect at least in some circumstances. This is an essentially manipulative condition that could be immediately reformulated in terms of counterfactuals (interventions).

Given the above condition, we can safely assume that X is a *direct cause* of Y in a causal model M if and only if X appears on the right hand side of the equation for Y in M (cf. (Woodward 2003; Weslake 2015)).

The relation of direct causation can be depicted graphically as a *causal diagram* of a causal model. As has been demonstrated, e.g., in (Pearl 2000), such a diagram provides important information for causal reasoning in the structural account.

Now we are going to reformulate the above ideas and constructs in the logical setting of the causal calculus.

To begin with, we will restrict our attention to *clausal* causal theories that include only causal rules of the form $L \Rightarrow l$, where l is a literal, and L a set of literals.

We will assume that *each* causal rule of a causal theory represents an autonomous causal mechanism. This assumption presupposes, however, that the causal theory does not contain redundant causal rules that are logically subsumed by other rules. The following definition makes this requirement precise:

Definition 11. A causal theory Δ will be called *parsimonious* if no causal rule from Δ is derivable from the rest of the rules in Δ by causal inference.

Among other things, the above notion of parsimony secures that the antecedents of the causal rules for each literal in a causal theory jointly form a *minimal* necessary condition for the latter. This constraint constitutes one of the important amendments to the traditional regularity approach to causation (see (Baumgartner 2013)).

As in the structural approach, given the above restriction, we will identify direct causes as literals that appear in the bodies of the causal rules that cause the effect.

Definition 12. A literal l_0 will be said to be a *direct cause* of a literal l in a clausal causal theory Δ if Δ contains a rule of the form $l_0, L \Rightarrow l$.

The following theorem shows that the above notion of a direct cause can also be given a counterfactual description.

Theorem 5. *l_0 is a direct cause of l in a parsimonious Pearl causal theory Δ if and only if there exists a set L of literals built from the rest of the atoms in Δ such that $L, l_0 >_\Delta l$ and $L, \neg l_0 >_\Delta \neg l$.*

The above condition can be viewed as a logical counterpart of the test pair condition in the structural account. Just as the latter condition, it secures that each literal in the body of a causal rule is a difference-maker for its head (cf. (Baumgartner 2015)). At the same time, the above theorem also shows that counterfactuals are expressive enough to capture the notion of a direct cause and causal diagram of a causal theory.

Intervention-Equivalence and Basic Inference

According to (Pearl 2000), every causal model stands not for just one but for a whole set of its submodels that embody interventional contingencies. These submodels determine the 'causal content' of a given causal model. In accordance with this, the following definition has been introduced (though in a different terminology) in (Bochman and Lifschitz 2015):

Definition 13. Determinate causal theories Γ and Δ are *intervention-equivalent* (*i-equivalent*, for short) if, for every set L of literals, the revision $\Gamma * L$ has the same nonmonotonic semantics as the revision $\Delta * L$.

Now, under general finiteness restrictions that have been adopted in this study, it is easy to show that intervention-equivalence of two causal theories amounts to coincidence of their associated counterfactuals:

Theorem 6. *Determinate causal theories Δ and Γ are i-equivalent iff they determine the same counterfactuals: for any set L of literals, and any A,*

$$L >_\Delta A \text{ iff } L >_\Gamma A.$$

Thus, i-equivalence provides a useful tool for the study of causal counterfactuals and their expressive capabilities.

Remark. In (Bochman and Lifschitz 2015), an attempt has been made to connect intervention equivalence with the causal equivalence in the causal calculus. The results, however, were not entirely satisfactory. Thus, though it has been shown that intervention equivalence implies equivalence with respect to causal inference, the reverse implication has been shown to hold only for a very narrow class of *modular* causal theories. In contrast, in this study we are going to show that intervention equivalence is intimately connected with a stronger equivalence with respect to basic causal inference.

To begin with, our next result will show that interventions cannot distinguish causal theories that are basically equivalent:

Theorem 7. *Basically equivalent determinate causal theories are intervention equivalent.*

Furthermore, being combined with our preceding result, the above theorem implies, in effect, that the language of counterfactuals does not allow to distinguish basically equivalent sets of causal rules.

It has been shown in (Bochman and Lifschitz 2015) (using a suitable counterexample) that causal equivalence does not imply intervention-equivalence. In other words, there are causal theories that are causally equivalent, but their revisions with the same literal determine different causal worlds (and counterfactuals). Strengthening this result, the following example shows that even regular equivalence does not imply intervention equivalence.

Example 2. The causal theories

$$\Delta = \{p \Rightarrow q, p \wedge q \Rightarrow r\}$$

and

$$\Gamma = \{p \Rightarrow q, p \Rightarrow r\}$$

are regularly equivalent. However, their respective revisions $\Delta * \neg q = \{\mathbf{t} \Rightarrow \neg q, p \wedge q \Rightarrow r\}$ and $\Gamma * \neg q = \{\mathbf{t} \Rightarrow \neg q, p \Rightarrow r\}$ are already not regularly equivalent. Moreover, if p is exogenous, they have different causal worlds: $\{p, \neg q, r\}$ is a causal world of $\Gamma * \neg q$, but not of $\Delta * \neg q$.

Theorem 7 above implies that the set of counterfactuals that hold with respect to a causal theory cannot determine uniquely the source causal theory, already because it cannot distinguish basically equivalent theories. Still, the following theorem will show that, up to the basic equivalence, there is indeed a one-to-one correspondence between Pearl causal theories and their associated counterfactuals.

Theorem 8. *If Δ is a Pearl causal theory, then, for any literal l and any set L of literals such that $l \notin L$,*

$$L \Rightarrow^b_\Delta l \text{ if and only if } L' >_\Delta l,$$

for any set of literals $L' \supseteq L$ such that $\neg l \notin L'$.

The above result implies that the set of counterfactuals that hold with respect to a causal theory uniquely determines the 'basic closure' ($\Rightarrow^b_\Delta$) of the latter. As an immediate consequence, we obtain

Corollary 9. *Pearl causal theories are intervention equivalent iff they are basically equivalent.*

It should be noted, however, that the above correspondence cannot be extended to arbitrary causal theories. In fact, the following counterexample, given in (Bochman and Lifschitz 2015), can also be used in the present context:

Example 3. Causal theories $\{p \Rightarrow p\}$ and $\{\mathbf{t} \Rightarrow p\}$ are not equivalent even for causal inference relations. Still, it is easy to verify that they are intervention equivalent, since all their possible revisions have the same causal worlds.

Basic equivalence in the structural account. As we have mentioned in the Introduction, it has often been argued that structural equation models and counterfactuals are essentially equivalent formalisms (see, e.g., (Galles and Pearl

1998), (Hitchcock 2007)). The above results can now be used to justify this claim.

Our last result above makes basic inference an 'internal' causal logic of interventions and counterfactuals in Pearl causal theories. It should be noted, however, that basic equivalence (i.e., the validity of the Or rule of inference) is in some sense 'built-in' in Pearl's structural account of causation itself due to the underlying assumption that any endogenous variable is determined by a *single* causal mechanism (formulated as a structural equation). Indeed, according to this principle, all the alternative causal factors that determine a given (Boolean) endogenous variable should be conjoined by disjunction into a single formula by the very definition of the structural equation. Consequently, basically equivalent causal descriptions are indistinguishable in the language of structural equations. But then the above results will imply, in effect, that there is indeed a one to one correspondence between Pearl's causal models and their associated sets of counterfactuals.

As we are going to show, however, the assumption of a single mechanism, and the ensuing 'collapse' of basically equivalent causal descriptions could be viewed as the main source of the problem of structural equivalents in the structural accounts of actual causation.

Counterfactuals and Actual Causation

Actual causation (aka 'singular causation', or 'causation in fact') deals with causal claims of the form *"C was a cause of E"*. In other words, it deals with *post factum* attribution of causal responsibility for actual outcome. Following (Lewis 1973), an overwhelming majority of current approaches to this notion attempt to define it in terms of counterfactuals.

The starting point of all counterfactual accounts is the but-for test (*sine qua non*) commonly used in both tort law and criminal law. The test asks, "but for C, would E have occurred?" If the answer is yes, then C is an actual cause of E. However, taken as a counterfactual assertion, the but-for test breaks down in cases of redundant causation (e.g., preemption or overdetermination), wherefore, using David Lewis' phrase, we need extra bells and whistles.

A broad scheme of generalizing the but-for test can be described using the notion of *"de facto dependence"* from (Yablo 2004): E de facto depends on C just in case had C not occurred, and had other suitably chosen factors been held fixed, then E would not have occurred. The trick is to say what "suitably chosen" means. The majority of counterfactual approaches, including the prominent HP definitions of Halpern and Pearl (Halpern and Pearl 2005; Halpern 2016a), could be viewed as particular instantiations of this general scheme. In many accounts, the "suitably chosen" parameters are determined by a path of counterfactual dependencies that should exist between the cause and effect (see, e.g., (Woodward 2003; Hitchcock 2007; 2001; Weslake 2015)), though the HP definitions are more general.

It is extremely difficult to adjudicate the advantages and shortcomings of the host of counterfactual accounts that have been suggested in the literature and, even more generally, the precise role of counterfactuals in assertions of actual causation (cf. (Menzies 2011)). We will argue, however, that there are some general, 'blanket' problems for the counterfactual approach to actual causation that transcend the boundaries of specific definitions.

To begin with, beyond the but-for test, there are no general principles, or 'rationality postulates', for the choice of the 'right' counterfactual definition of actual causation. This creates an obvious trust problem for any potential definition of this kind[3]. In practice, most of this research is largely example-driven, but even on this score there are grave doubts whether the current 'empirical pool' of examples is sufficiently representative (see (Glymour et al. 2010)).

In an important, though often overlooked, paper (Maudlin 2004), Tim Maudlin has forcefully argued that there are no direct analytical connections between (actual) causation and counterfactuals in either direction. In his first 'thought experiment', he suggested to consider a world in which all forces are extremely short range (within an angstrom), and there is a particle P that is at rest at t_0 and moving at t_1, and that in the period between t_0 and t_1 only one particle, particle Q, came within an angstrom of P. Then we know with complete certainty what caused P to start moving: It was the collision with Q. As Maudlin has argued, once we know the laws, we can make this causal claim without being certain about the validity of any associated counterfactual. And indeed, as we will see in the next section, the well-known INUS condition can be used to provide a natural definition of actual causation that does not use counterfactuals.

Maudlin's second example was purported to show that fixing truth values for all counterfactuals does not always fix the truth values of all causal claims. This example will turn out to be intimately related to our preceding results.

Example 4 (**Game of Life**). John Conway's Game of Life is played on a square grid. At any moment, each square in the grid is either empty or occupied, which depends on the how that square and the eight immediately adjacent squares were occupied at the previous moment. The rules of the game cover all possibilities, namely they specify for each of the 512 possible patterns of occupation of a 3-by-3 grid whether the central square is or is not occupied at the next instant. Consequently, the state of the grid evolves deterministically through time. Moreover, the rules thereby determine a unique truth value for every counterfactual assertion about this game.

Let us introduce some logical notation. Assume some fixed enumeration of the nine squares of a 3-by-3 grid, 0 being the central square, and let $p_i, i = 0, \ldots, 8$ denote the fact that the i-th square is occupied, while l_i will denote the corresponding literals. Then any rule of the Game of Life can be written as a causal rule of the form $A \Rightarrow l_0$, where A is a propositional formula in this language.

Suppose now that there are two patterns of occupation that differ only on square 1, but which both yield that the central square 0 will be occupied. There are, however, two possibil-

[3] "Consider some other relation, *schmausation*, which can be defined in terms of counterfactual dependence, adding drums and trumpets instead of bells and whistles. From the perspective of the counterfactual theory, schmausation is no less natural or distinctive." (Hitchcock 2011)

ities about how these transitions are generated by the rules. One possibility is that there is a *single* mechanism behind both these transitions which does not involve square 1; this mechanism can be encoded, for instance, by a causal rule of the form

$$l_2, \ldots, l_8 \Rightarrow p_0$$

Another possibility, however, is that these two transitions are governed by two *different* mechanisms, so they are instantiations of two different laws, for instance

$$p_1, l_2, \ldots, l_8 \Rightarrow p_0 \quad \text{and} \quad \neg p_1, l_2, \ldots, l_8 \Rightarrow p_0.$$

The above difference does not affect the associated counterfactuals, but it influences our causal judgments. Thus, for a transition from a pattern where square 1 is occupied, p_1 can be naturally viewed as one of the causes of p_0 for the case of alternative mechanisms, though not in the case of a single mechanism. As has been noted by Maudlin, where the laws are in dispute, the *causes* are in dispute, all while the truth values of the counterfactuals remain unquestioned.

As can be seen, the above two possibilities correspond to two basically equivalent causal theories, so our previous results determine that these theories are indistinguishable by counterfactual means. However, the main lesson from the above example is that such theories may still support different claims about actual causation.

It turns out that the above phenomenon could also be held responsible for the problem of 'structural equivalents' in the counterfactual approaches to actual causation. It can be illustrated on the following preemption example from (Pearl 2000):

Example 5 (**Desert Traveler**). Enemy 1 poisons T's canteen (p), and enemy 2, unaware of enemy 1's action, shoots and empties the canteen (x). A week later, T is found dead (y).

An enriched causal model, appeared in (Pearl 2000), included also variables C (for cyanide intake) and D (for dehydration) and contained the following equations:

$$c = p \wedge \neg x \quad d = x \quad y = c \vee d$$

If we substitute c and d into the expression for y, we obtain a disjunction

$$y = x \vee (p \wedge \neg x)$$

Pearl has argued, however, that, though $x \vee (\neg x \wedge p)$ is logically equivalent to $x \vee p$, these two expressions are not 'structurally equivalent', and it is this asymmetry that makes us proclaim x and not p to be the cause of death.

There is a clear 'anti-logical' overtone in the above Pearl's argument that seems to suggest that purely logical descriptions do not always provide an adequate representation of the relevant 'structural' differences. We suggest, however, a somewhat different diagnosis. As we have mentioned, basic equivalence is built in Pearl's structural account due to the underlying assumption that any endogenous variable is determined by a single causal mechanism. Recall in this respect that the example of Maudlin was based on a possibility that a variable (square occupation) is determined by two different, *alternative*, mechanisms. In the structural account, however, all such mechanisms are combined by disjunction into a single equation, so the relevant structural differences could be preserved only if we either sacrifice logical equivalence, or use auxiliary variables (C and D in the above example).[4] The regularity approach that we will describe in the next section, allows for a more succinct description of the above example without sacrificing classical equivalence.

Actual Causation on the Regularity Approach

The regularity approach originates in the 'covering law' analysis of causation by J. S. Mill, and is based on the well-known INUS condition of (Mackie 1974). A perspicuous formulation has been given in (Wright 1985):

> The NESS[5] test: a condition c was a cause of a consequence e if and only if it was necessary for the sufficiency of a set of *existing* antecedent conditions that was sufficient for the occurrence of e.

Modern regularity theories (Baumgartner 2008; Grasshoff and May 2001) have successfully met some of the traditional challenges to the original theory by adopting more stringent conditions on necessary and sufficient conditions (see (Baumgartner 2013) for an overview). However, a more radical amendment has been suggested in (Strevens 2007), according to which the very notion of sufficiency (which has been assumed to be classical in the original regularity theory) should be given a causal interpretation. This view has been endorsed by the author of the NESS test himself:

> The required sense of sufficiency, which I call 'causal sufficiency' to distinguish it from mere lawful strong sufficiency, is the instantiation of all the conditions in the antecedent ('if' part) of a causal law, the consequent ('then' part) of which is instantiated by the consequence at issue. (Wright 2013)

According to Wright, a sequence of such causal laws that links the condition at issue with the consequence will provide the required justification for the causal claim.

The definition of actual causation in (Bochman 2018) has been based on an explication of the relevant notion of causal sufficiency in terms of causal inference.

An actual causation claim presupposes a given causal theory Δ, and an actual world α that is a causal (exact) world with respect to Δ. For reasons explained in (Bochman 2018), Δ is required to be a parsimonious causal theory.

Definition 14. Let Δ be a clausal causal theory, and α a causal world of Δ.

- A causal rule $l_1, \ldots, l_n \Rightarrow l$ will be called *active in* α if $\{l_1, \ldots, l_n\} \subseteq \alpha$.
- The *actual sub-theory* of Δ wrt α is the set of all causal rules from Δ that are active in α.

[4] Introduction of auxiliary variables has even been suggested as a modelling rule in (Halpern and Pearl 2005): "If we want to argue in a case of preemption that c is a cause of e rather than d, then there must be a random variable ... that takes on different values depending on whether c or d is the actual cause."

[5] Necessary Element of a Sufficient Set

Remark. There is a lot of similarity, both in content and purpose, between the actual causal sub-theory and the notion of a *causal beam* from Chapter 10 of (Pearl 2000). Moreover, our parsimony restriction to non-redundant rules naturally corresponds to Pearl's minimality requirement on functions $\{f_i\}$ in a causal model that should not contain redundant variables. All this makes our definition below much similar to the original definition of actual causation, described in (Pearl 2000). It should be noted, however, that our construction is immune to the counterexamples that have led to abandoning this idea by Pearl.

Let Δ_α denote the actual sub-theory of Δ wrt α, while $\Rightarrow_\alpha$ will denote the associated causal inference (i.e., $\Rightarrow_{\Delta_\alpha}$).

Definition 15 (actual cause)**.** Let α be a causal world of a parsimonious causal theory Δ. A literal $l_0 \in \alpha$ will be said to be an *actual cause* of a literal l in α wrt Δ if and only if there exists a set of literals $L \subseteq \alpha$ such that

1. $l_0, L \Rightarrow_\alpha l$;
2. $L \not\Rightarrow_\alpha l$.

It has been shown in (Bochman 2018) that causal inference with respect to the actual sub-theory $\Rightarrow_\alpha$ can be replaced with an unconstrained *regular* inference $\Rightarrow^r_\Delta$.

Corollary 10. *Let α be a causal world of a clausal causal theory Δ. Then $l_0 \in \alpha$ is an actual cause of l in α wrt Δ if and only if there exists a set of literals $L \subseteq \alpha$ such that*

1. $l_0, L \Rightarrow^r_\Delta l$;
2. $L \not\Rightarrow^r_\Delta l$.

The above description makes our definition of actual causation a straightforward formalization of the NESS test with regular inference as a logical explication of causal sufficiency. As a consequence, regularly equivalent causal theories support the same claims of actual causation.

Example 6 (**Desert Traveler**, revisited). The causal model for the Desert Traveler story can be translated into the following causal theory:

$$p \wedge \neg x \Rightarrow c \quad x \Rightarrow d \quad c \Rightarrow y \quad d \Rightarrow y$$
$$\neg p \Rightarrow \neg c \quad x \Rightarrow \neg c \quad \neg x \Rightarrow \neg d \quad \neg c, \neg d \Rightarrow \neg y$$

The actual world is $\{p, x, y, \neg c, d\}$, so the actual sub-theory is

$$x \Rightarrow d \quad d \Rightarrow y \quad x \Rightarrow \neg c$$

Accordingly, shot (x) and dehydration (d), but not poison (p), are actual causes of death (y).

In our logical framework, the asymmetry between the preempting and preempted cause stems from the fact that basically equivalent sets of causal rules might be regularly non-equivalent, so they could support different assertions of actual causation. In the present case, the causal theory $\{x \Rightarrow c, \neg x \wedge p \Rightarrow c\}$ is not regularly equivalent to $\{x \Rightarrow c,\ p \Rightarrow c\}$, though they are equivalent with respect to basic inference. Furthermore, in contrast to the structural account, the auxiliary variables c and d are not necessary for describing this example; the following simple causal theory provides the same answers about actual causation among the salient variables $\{x, p, y\}$:

$$p \wedge \neg x \Rightarrow y \quad x \Rightarrow y \quad \neg x, \neg p \Rightarrow \neg y$$

Conclusions

We have provided a formal definition of counterfactuals in the causal calculus, which has given us an opportunity to investigate their expressive capabilities in describing causation. It has been shown, in particular, that the counterfactual language has essential logical limitations (compared with causal rules), namely it obliterates distinctions between basically equivalent causal theories. However, such theories can provide different answers about actual causation, which could be seen as the main source of the problem of structural equivalents in counterfactual approaches to causation.

The above results have obvious implications for the trilemma of relations between causation, laws and counterfactuals. Basically, we believe that these results demonstrate that, contrary to the currently dominant opinions, counterfactuals (at least on their standard understanding) cannot serve as a *ground* neither for (causal) laws, nor even for actual causation.

In (Maudlin 2004), Tim Maudlin has argued that (actual) causation and counterfactuals are analytically independent notions, whereas the correlations between them are due to the common "third factor", namely natural laws and law-like regularities that provide an ultimate basis for both. This view is remarkably close to the suggested representation of causality in the causal calculus, since in the latter causal rules serve as causal laws that provide an ultimate basis for causal reasoning, including both counterfactuals and actual causation. Moreover, in full accordance with Maudlin's views, our respective definitions of these notions do not have direct analytical connections with each other, though both are formulated in terms of causal rules.

The above general picture has much in common also with the original structural account of (Pearl 2000) that uses structural equations as a primary causal formalism. The structural account assigns, however, a paramount role to interventions and counterfactuals in causal reasoning, and the above considerations should not be construed as an argument against this role. They suggest, however, that the relations between counterfactuals and causation are less straightforward than what has been usually thought, especially for actual causation. We have mentioned also that the basic equivalence is in a sense built in the structural account itself due to the assumption that each structural equation describes a single causal mechanism. This creates situation in which introduction of auxiliary variables is apparently the only way to distinguish logically equivalent conditions that are not 'structurally' equivalent, whether we use a counterfactual approach or not. This produces, in turn, a seemingly problematic dimension of variability, or instability, of causal claims depending on the auxiliary variables we use (see (Halpern 2016b)). In contrast, in our logical approach, each causal rule of a parsimonious causal theory is viewed as representing an independent causal mechanism, which produces more adequate descriptions for problematic cases of actual causation. At least for propositional (Boolean) variables, causal rules seem to provide clear representational advantages over both counterfactuals and structural equations.

References

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

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

Baumgartner, M. 2015. Parsimony and causality. *Quality & Quantity* 49(2):839–856.

Bochman, A., and Lifschitz, V. 2015. Pearl's causality in a logical setting. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, 1446–1452. AAAI Press.

Bochman, A. 2003. A logic for causal reasoning. In *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, 141–146. Acapulco: Morgan Kaufmann.

Bochman, A. 2004. A causal approach to nonmonotonic reasoning. *Artificial Intelligence* 160:105–143.

Bochman, A. 2018. Actual causality in a logical setting. In *Proceedings 27th Int. Joint Conf. on Artificial Intelligence and the 23rd European Conf. on Artificial Intelligence, IJCAI-ECAI-18*.

Galles, D., and Pearl, J. 1998. An axiomatic characterization of causal counterfactuals. *Foundations of Science* 3(1):151–182.

Giunchiglia, E.; Lee, J.; Lifschitz, V.; McCain, N.; and Turner, H. 2004. Nonmonotonic causal theories. *Artificial Intelligence* 153:49–104.

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

Grasshoff, G., and May, M. 2001. Causal regularities. In Spohn, W.; Ledwig, M.; and Esfeld, M., eds., *Current issues in causation*. Paderborn: Mentis. 85–114.

Halpern, J. Y., and Pearl, J. 2001. Causes and explanations: A structural-model approach—part I: Causes. In *Proc. Seventh Conf. On Uncertainty in Artificial Intelligence (UAI'01)*, 194–202. San Francisco, CA: Morgan Kaufmann.

Halpern, J. Y., and Pearl, J. 2005. Causes and explanations: A structural-model approach. part I: Causes. *British Journal for Philosophy of Science* 56(4):843–887.

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

Halpern, J. Y. 2016b. Appropriate causal models and the stability of causation. *Rew. Symb. Logic* 9(1):76–102.

Hitchcock, C. 2001. The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy* XCVIII(6):273–299.

Hitchcock, C. 2007. Prevention, preemption, and the principle of sufficient reason. *Philosophical Review* 116:495–532.

Hitchcock, C. 2011. The metaphysical bases of liability: Commentary on Michael Moore's "Causation and Responsibility". *Rutgers Law Journal* 42(2):377–404.

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

Lifschitz, V. 1997. On the logic of causal explanation. *Artificial Intelligence* 96:451–465.

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

Makinson, D., and van der Torre, L. 2000. Input/Output logics. *Journal of Philosophical Logic* 29:383–408.

Maudlin, T. 2004. Causation, counterfactuals, and the third factor. In Collins, J.; Hall, N.; and Paul, L. A., eds., *Counterfactuals and Causation*. Cambridge MA: MIT Press.

McCain, N., and Turner, H. 1997. Causal theories of action and change. In *Proceedings AAAI-97*, 460–465.

Menzies, P. 2011. The role of counterfactual dependence in causal judgements. In Hoerl, C.; McCormack, T.; and Beck, S. R., eds., *Understanding counterfactuals, understanding causation*. Oxford: Oxford University Press.

Pearl, J. 2000. *Causality: Models, Reasoninig and Inference*. Cambridge UP. 2nd ed., 2009.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT press, 2nd edition.

Strevens, M. 2007. Mackie remixed. In Campbell, J. K.; O'Rourke, M.; and Silverstein, H. S., eds., *Causation and Explanation*. Cambridge MA: MIT Press.

Weslake, B. 2015. A partial theory of actual causation. *British Journal for the Philosophy of Science*. To appear.

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

Wright, R. W. 1985. Causation in tort law. *California Law Review* 1735:1788–91.

Wright, R. W. 2013. The NESS account of natural causation: A response to criticisms. In Kahmen, B., and Stepanians, M., eds., *Causation and Responsibility: Critical Essays*. De Gruyter.

Yablo, S. 2004. Advertisement for a sketch of an outline of a proto-theory of causation. In Collins, J., N. H., and Paul, L. A., eds., *Causation and Counterfactuals*. Cambridge: MIT Press. 119–138.