Probabilistic Definitions of Actual Causation Using CP-logic

Sander Beckers and Joost Vennekens
KULeuven - University of Leuven
{Sander.Beckers,Joost.Vennekens}@cs.kuleuven.be

Abstract
Since Pearl’s seminal work on providing a formal language for causality, the subject has garnered a lot of interest among philosophers and researchers in artificial intelligence alike. One of the most debated topics in this context regards the notion of actual causation, which concerns itself with specific – as opposed to general – causal claims. The search for a proper formal definition of actual causation has evolved into a controversial debate, that is pervaded with ambiguities and confusion. The goal of our research is twofold. First, we wish to provide a clear way to compare competing definitions. Second, we also want to improve upon these definitions so they can be applied to a more diverse range of instances, including non-deterministic ones. To achieve these goals we will provide a general, abstract definition of actual causation, formulated in the context of the expressive language of CP-logic (Causal Probabilistic logic). We will then show that three recent definitions by Ned Hall (originally formulated for structural models) and a definition by Beckers and Vennekens (formulated for CP-logic directly) can be viewed and directly compared as instantiations of this abstract definition, which allows them to deal with a broader range of examples.

1 Introduction
Suppose we know the causal laws that govern some domain, and that we then observe a story that takes place in this domain; when should we now say that, in this particular story, one thing actually caused another? Ever since Lewis (1973) first analyzed this problem of actual causation (a.k.a. token causation) in terms of counterfactual dependence, philosophers and researchers from the AI community alike have been trying to improve on his attempt. Following the work of Pearl (2000), structural equations have recently become a popular formal framework for this (Hitchcock 2007; 2009; Halpern and Pearl 2005a; Halpern and Hitchcock). A notable exception is the work of Ned Hall, who has extensively critizised the privileged role of structural equations for causal modelling, as well as the definitions that have been expressed with it (Hall 2007). He has proposed several definitions himself (Hall and Paul 2003; Hall 2004; 2007), the latest of which is a sophisticated attempt to overcome the flaws he observes in those that rely too heavily on structural equations. Another approach is taken by Beckers and Vennekens, who develop a definition of their own within the framework of CP-logic (Causal Probabilistic logic) (Beckers and Vennekens 2012; Vennekens 2011).

The standard procedure to defend one’s definition is to go over certain pivotal examples and argue that the proposed definition matches up with basic intuitions regarding these. If, however, one is confronted with an example that gives rise to stubborn counterintuitive answers, the strategy is to argue that the usual way of portraying the example is misleading, and that when modelled properly - i.e., using ones own formal resources - these counterintuitions disappear. A more conceptual discussion based on the underlying motivations, rather than merely keeping a tally, is hindered by the fact that the different definitions are expressed using different frameworks, such as neuron diagrams, structural equations, CP-logic, or simply in natural language.

The aim of our research paper is to remove this obstacle, by illustrating how the expressive framework of CP-logic can serve as a general language for expressing and comparing different definitions of actual causation. Further, we also want to improve upon these definitions so they can be applied to a more diverse range of instances, including non-deterministic ones. To achieve these goals we will provide a general, abstract definition of actual causation. We then show how both the definition of Beckers and Vennekens, and the three definitions by Ned Hall, can be viewed and directly compared as instantiations of this abstract definition, which allows them to deal with a broader range of examples.

2 CP-logic
We give a short, informal introduction to CP-logic. A detailed description can be found in (Vennekens, Denecker, and Bruynooghe 2009; 2010). The basic syntactical unit of CP-logic is a CP-law, which takes the general form of $\text{Head} \leftarrow \text{Body}$. The body can in general consist of any first-order logic formula. However, in this paper, we restrict our attention to propositional formulas in CNF. The head contains a disjunction annotated with probabilities, representing
the possible effects of this law. These effects are all atoms, with the possible exception of one disjunct, which can remain empty. The latter represents the fact that the law may not have an effect at all.

Simply put, each CP-law models a specific causal mechanism. Informally, we take this to mean that, if the Body of the law is satisfied, then this will cause precisely one of the disjuncts that figure in the Head, each with their respective probabilities. If a disjunct is chosen containing an atom, then this atom becomes true (unless of course it’s already true); if the empty disjunct is chosen, the law has no effect. Thus CP-logic is a dynamic language, where laws are applied one after another. When a law is applied, i.e., a disjunct is chosen and possibly an atom becomes true, we will say that this law happens. A set of such CP-laws forms a CP-theory, and represents the causal structure of the domain at hand. We illustrate with an example from (Hall 2004):

Suzy and Billy each decide to throw a rock at a bottle. When Suzy does so, her rock shatters the bottle with probability 0.9. Billy’s aim is slightly worse and he only hits with probability 0.8.

This small causal domain can be expressed by the following CP-theory $T$:

\[
\text{Throws}(Suzy) \leftarrow . \\
\text{Throws}(Billy) \leftarrow . \\
(\text{Breaks} : 0.9) \lor (\text{true} : 0.1) \leftarrow \text{Throws}(Suzy). \\
(\text{Breaks} : 0.8) \lor (\text{true} : 0.2) \leftarrow \text{Throws}(Billy).
\]

(1) (2) (3) (4)

This causal model consists of four causal laws. The first two are both vacuous (i.e., they will happen in every story) and deterministic (i.e., they have only one possible outcome). The last two laws are non-deterministic, causing either the bottle to break or nothing at all. From now on, we will leave the empty disjunct true implicit.

The given theory summarizes all possible stories that can take place in this model. For example, it allows for the story consisting of the following chain of events:

Suzy and Billy both throw a rock at a bottle. Suzy’s rock gets there first, shattering the bottle. Since both throws are perfectly accurate, Billy’s would have shattered the bottle had it not been preempted by Suzy’s throw.

To formalize this idea, the semantics of CP-logic uses probability trees (Shafer 1996). For this example, one such tree is shown in Figure 1. Here, each node represents a state of the domain, which is characterized by an assignment of truth values to the atomic formulas, in this case $\text{Throws}(Suzy)$, $\text{Throws}(Billy)$ and $\text{Breaks}$. In the initial state of the domain (the root node), all atoms are assigned their default value false. In this example, the bottle is initially unbroken and the rocks are still in Billy and Suzy’s hands. The children of a node $x$ are the result of a law happening: each edge $(x, y)$ corresponds to a specific disjunct that was chosen from the head of the law that was applied in node $x$. In this case law (1) is applied first, so the assignment in the child-node is obtained by setting $\text{Throws}(Suzy)$ to true, its deviant value. The third state has two child-nodes, corresponding to law (3) being applied and either breaking the bottle (left child) or not (right child). The leftmost branch is thus the formal counterpart of the above story, where the last edge represents the fact that Billy’s throw was also accurate, even though there was no bottle left to break. A branch ends when no more laws can be applied.

It’s important to note that any causal law can happen only once, and will cause precisely one of its disjuncts. Also, a causal law can only happen in a state whose assignment of truth values satisfies the law’s precondition. However, this condition is not enough for preconditions containing a negated atom, as this may lead to ambiguities in the semantics. This problem can be solved by requiring that the relevant atom should not only be currently assigned false, but that we also have to be certain that it will remain so. In other words, the atom should actually have become impossible. The formal semantics of CP-logic uses ideas from Kleene’s three-valued logic to define this mathematically.

A probability tree corresponding to a theory $T$ in CP-logic defines an a priori probability distribution $P_T$ over all things that might happen in this domain, which can be read off the leaf nodes. For instance, the probability of the bottle breaking is the sum of the probabilities of the leaves in which Breaks is true – the white circles in Figure 1 – giving 0.98. We have shown here only one such probability tree, but we can construct others as well by applying the laws in different orders. An important property however is that all trees defined by the same theory result in the same probability distribution.

2.1 Some Terminology

We specify some operations on CP-logic theories that will be used throughout this paper. Assume we have a theory $T$ and a branch $b$ of one of $T$’s probability trees, such that both $C$ and $E$ hold in its leaf. To make $T$ deterministic in accordance with the choices made in $b$, we transform $T$ into $T^b$ by replacing the heads of the laws that happened in $b$ with the atoms which were chosen from those heads in $b$. For example, if we take as branch $b$ the previous story regarding
Suzy and Billy, then $T^b$ would be:

$$\text{Throws}(\text{Suzy}) \leftarrow \cdot.$$  
(5)

$$\text{Throws}(\text{Billy}) \leftarrow \cdot.$$  
(6)

$$\text{Breaks} \leftarrow \text{Throws}(\text{Suzy}).$$  
(7)

$$\text{Breaks} \leftarrow \text{Throws}(\text{Billy}).$$  
(8)

We will use Pearl’s $do()$-operator to indicate an intervention (Pearl 2000). The intervention on a theory $T$ that makes variable $C$ false, denoted by $do(\neg C)$, removes $C$ from the head of any law in which it occurs, yielding $T|do(\neg C)$. For example, to prevent Suzy from throwing, the resulting theory $T|do(\neg \text{Throws(Suzy)})$ is given by:

$$T \leftarrow \cdot.$$  
(9)

$$\text{Throws}(\text{Billy}) \leftarrow \cdot.$$  
(10)

$$(\text{Breaks : 0.9}) \lor (\text{true : 0.1}) \leftarrow \text{Throws}(\text{Suzy}).$$  
(11)

$$(\text{Breaks : 0.8}) \lor (\text{true : 0.2}) \leftarrow \text{Throws}(\text{Billy}).$$  
(12)

Laws with an empty head, such as (9), can also simply be omitted. The analogous operation $do(C)$ on a theory $T$ corresponds to adding the deterministic law $C \leftarrow \cdot$.

### 3 Defining Actual Causation Using CP-logic

In this section, we formulate a general definition of actual causation, which can accommodate a wide range of concrete definitions by filling in details that we first leave open. We demonstrate this by looking at definitions by Hall and by Beckers and Vennekens.

#### 3.1 Actual Causation in General

We first make precise the notion of counterfactual dependence. As Pearl showed in (Pearl 2000), in the context of structural equations counterfactuals can be obtained as syntactic transformations of a causal model. The same goes for CP-logic, as was shown in (Vennekens, Denecker, and Bruynooghe 2010), by using the $do(\cdot)$ operation introduced in the previous section.

For reasons of simplicity, Hall, along with the majority of approaches, only considers deterministic causal relations. Further, it is taken for granted that the actual values of all variables are given. In such a context, counterfactual dependence of the event $E$ on $C$ is expressed by the conditional: if $do(\neg C)$ then $\neg E$, where it is assumed that all exogenous variables take on their actual values. In our probabilistic setting, the latter translates into making those laws that actually happened deterministic in accordance with the choices made in the story. However in many cases some exogenous variables simply do not have an actual value to start with. For example, if Suzy is prevented from throwing her rock, then we cannot say what the accuracy would have been had she done so. In CP-logic, this would be represented by the fact that law (3) is not applied. Hence, in a more general setting, it is required only that $do(\neg C)$ makes $\neg E$ possible.

**Definition 1.** Given a theory $T$, we say that $E$ is counterfactually dependent on $C$ according to $T$ iff $P_T(\neg E|do(\neg C)) > 0$.

This definition is the CP-logic equivalent of the one given in (Halpern and Pearl 2005b, p. 27), where they discuss probabilistic causal models. When dealing with actual causation, however, the question of counterfactual dependence is posed in a more specific context, namely in relation to an actual story that took place. Thus we introduce the following definition.

**Definition 2.** Given a theory $T$ and a story $b$, we say that $E$ is counterfactually dependent on $C$ in $b$ iff $P_{T^b}(\neg E|do(\neg C)) > 0$.

This counterfactual dependence relation represents only one way of taking into account the actual story in which $C$ and $E$ took place. However, in general other options are available, given by specifying other ways to modify the general causal theory $T$ into a theory $T'$ that we derive from $T$, by taking into account the actual story in a suitable way. This comes down to removing laws from $T$ which played no part in the story, and making some of the laws in $T$ deterministic in accordance with the choices made in the story. We shall call these laws irrelevant – noted as $T_{Ir(b)}$ – respectively intrinsic – noted as $T_{In(b)}$ – with regards to a particular story. Different definitions of actual causation will define these two sets of laws in different ways. We construct $T'$ as $T \setminus (T_{Ir(b)} \cup T_{In(b)}) \cup T^b_{In(b)}$.

**Definition 3 (Actual causation).** Given a theory $T$ and a branch $b$ such that both $C$ and $E$ hold in its leaf. We define that $C$ is an actual cause of $E$ if and only if $E$ is counterfactually dependent on $C$ according to $T'$.

This definition is similar in spirit to that of a partial explanation given in (Halpern and Pearl 2005b). There the probability measures the goodness of the explanation, here it measures the strength of the cause.\(^1\)

#### 3.2 Beckers and Vennekens 2012 Definition

A recent proposal for a definition of actual causation was originally formulated in (Vennekens 2011), and later slightly modified in (Beckers and Vennekens 2012). Here, we summarize the basic ideas of the latter, and refer to it as BV12.

To follow the actual story as closely as possible, we force all laws that happened in $b$ to have the same effect as they had in $b$, even when considering counterfactual situations. In other words, all laws that happened in $b$ are intrinsic.

To decide which laws were relevant for causing $E$ in our story, we start from a simple temporal criterion: everything that happened after the effect $E$ took place is irrelevant, and everything that happened before isn’t. For example, to figure out why the bottle broke in our previous example, law (4) is considered irrelevant, because the bottle was already broken by the time Billy’s rock arrived. For laws that did not happen in $b$ at all, we distinguish laws that could still happen when $E$ occurred, from those that could not. The first are considered irrelevant, whereas the second aren’t. This ensures that any story $b'$ that is identical to $b$ up to and including the occurrence of $E$ provides the same judgements about the causes.

\(^1\)More recently, one of the authors of that paper has also extended the idea of gradation to the concept of causation, in (Halpern and Hitchcock ).
of $E$, since any law that did not happen in $b$ but does happen in $b'$, must obviously occur after $E$.

**BV12-Irrelevant.** A law $r$ of $T$ is irrelevant iff $r$ didn’t happen before $E$ in $b$, although it could have. (I.e., it was not impossible at the time when $E$ happened.)

**BV12-Intrinsic.** A law $r$ of $T$ is intrinsic iff $r$ happened in $b$.

### 3.3 Hall 2007

One of the currently most refined concepts of actual causation is that of (Hall 2007). Although Hall uses structural equations as a practical representational tool, he is of the opinion that intuitions about actual causation are best expressed using neuron diagrams. A key advantage of these diagrams, which they share with CP-logic, is that they distinguish between a default and deviant state of a variable. Proponents of structural equations, on the other hand, countered Hall’s approach by criticizing neuron diagrams’ limited expressivity (Hitchcock 2009, p. 398). Indeed, neuron diagrams are very limited in the kind of examples they can express, and therefore Hall is also limited in the kind of examples he can consider. In particular, neuron diagrams can only express deterministic causal relations and they also lack the ability to directly express *causation by omission*, i.e., that the absence of $C$ causes $E$. Hall’s solution is to argue against causation by omission altogether. By contrast, we will offer an improvement of Hall’s account that generalizes to a probabilistic context, and can also handle causation by omission. In short, we propose CP-logic as a way of overcoming the shortcomings of both structural equations and neuron diagrams.

**Neuron diagrams** We will shortly describe the functioning of a neuron diagram, and how this translates into CP-logic. A neuron can be in one of two states, one is the default “off” state and the other is the deviant “on” state in which the neuron “fires” or “is active”. Different kinds of links between two nodes define how the state of one affects the other. For instance, in (a), $E$ fires iff at least one of $B$ or $D$ fires, $D$ fires iff $C$ fires, and $B$ fires iff $A$ fires and $C$ doesn’t fire. Nodes that are “on” are represented by full circles and nodes that are “off” are shown as empty circles. So in (a), $A$, $C$, $D$ and $E$ all fire, whereas $B$ does not.

![Figure 2: (a)](image1)

![Figure 3: (b)](image2)

Diagrams (a) and (b) represent the same causal structure, but different stories: in both cases there are two causal chains leading to $E$, one starting with $C$ and another starting with $A$. But in (a) the chain through $B$ is preempted by $C$, whereas in (b) there is nothing for $C$ to preempt, as $A$ doesn’t even fire. Therefore (a) is an example of what is generally known as Early Preemption, whereas (b) is not.

Although Hall presents his arguments by using neuron diagrams, his definitions are formulated in terms of structural equations that correspond to such diagrams. This correspondence is quite straightforward: for each endogenous variable there is one equation, which contains a propositional formula on the right expressing the dependencies of the diagram in a concise way.

Any structural model $M$ can also be read as a CP-logic theory $T$. The firing of the neurons, and the resulting assignment to the variables in $M$, then correspond to a story $b$. One important difference between structural equations and CP-laws, is that we are not limited to using a single CP-law for each variable. As each law represents a separate causal mechanism, initiated by a different set of states of the neurons, we will represent dependencies such as that of $E$ by three laws, rather than one, corresponding to the three different ways in which $B$ and $D$ can cause $E$: each by themselves, or the two of them simultaneously. As a consequence, there are no disjunctions in the bodies of CP-laws representing neuron diagrams. The translation of examples (a) and (b) into CP-logic is given by the following CP-theory – where $p$ and $q$ represent some probabilities:

\[
\begin{align*}
(A : p) & \leftarrow . \\
(C : q) & \leftarrow . \\
B & \leftarrow A \land \neg C. \\
D & \leftarrow C.
\end{align*}
\]

**Hall 2007 Definition** The idea behind Hall’s definition is to check for counterfactual dependence in situations which are reductions of the actual situation, where a reduction is understood as “a variant of this situation in which strictly fewer events occur”. In other words, because the counterfactual dependence of $E$ on $C$ can be masked by the occurrence of events which are extrinsic to the actual causal process, we look at all possible scenario’s in which there are less of these extrinsic events. Hall puts it like this (2007, p. 129):

Suppose we have a causal model for some situation. The model consists of some equations, plus a specification of the actual values of the variables. Those values tell us how the situation actually unfolds. But the same system of equations can also represent nomologically possible variants: just change the values of one or more exogenous variables, and update the rest in accordance with the equations. A good model will thus be able to represent a range of variations on the actual situation. Some of these variations will be – or more accurately, will be modeled as – reductions of the actual situation, in that every variable will either have its actual value or its default value. Suppose the model has variables for events $C$ and $E$. Consider the conditional

\[
\text{if } C = 0; \text{ then } E = 0
\]

This conditional may be true; if so, $C$ is a cause of $E$. Suppose instead that it is false. Then $C$ is a cause of $E$ iff there is a reduction of the actual situation according to which $C$ and $E$ still occur, and in which this conditional is true.

Rather than speaking of fewer events occurring, in this definition Hall characterizes a reduction in terms of whether
or not variables retain their actual value. This is because in the context of neuron diagrams, an event is the firing of a neuron, which is represented by a variable taking on its deviant value, i.e., the variable becoming true. In the dynamic context of CP-logic, the formal object that corresponds most naturally to Hall’s informal concept of an event is the transition in a probability tree (i.e., the application of a causal law) that makes such a variable true. Therefore, in the context of CP-logic, we take a reduction to mean that no law is applied such that it makes a variable true that did not become true in the actual setting. To make this more precise, we introduce some new formal terminology. Throughout the following we take \( b \) and \( d \) to be branches of a probability tree of the theory \( T \) and assume that \( C \) and \( E \) both occur in \( b \).

**Definition 4.** Given a branch \( d \), the set of all laws that happened in \( d \) will be denoted by \( \text{Laws}_d \). The effect which was the result of the application of a rule \( r \in \text{Laws}_d \) — i.e., the disjunct of the head which was chosen — will be denoted by \( r_d \). If there was no effect — i.e., an empty disjunct was chosen — we say that \( r_d = 0 \). The set of true variables in the leaf of \( d \) will be denoted by \( \text{Leaf}_d \).

**Definition 5.** A branch \( d \) is a reduction of \( b \) iff \( \forall r \in \text{Laws}_d \colon \exists s \in \text{Laws}_b \colon r_d = s \lor r_d = 0 \).

Note that an equivalent definition of reduction is to state that \( \text{Leaf}_d \subseteq \text{Leaf}_b \). A reduction of \( b \) in which both \( C \) and \( E \) occur — i.e., hold in its leaf — will be called a \((C,E)\)-reduction. The set of all of these will be denoted by \( \text{Red}^{(C,E)}_d \). These reductions are precisely those branches which are relevant for Hall’s definition.

**Definition 6** (Hall’s definition in CP-logic). Given a theory \( T \) and a branch \( b \) such that both \( C \) and \( E \) hold in its leaf. We define that \( C \) is an actual cause of \( E \) if and only if \( (\exists d \in \text{Red}^{(C,E)}_b) : P_d(-E|do(-C)) > 0 \).

We will present a theorem to ensure the correctness of our translation.

**Theorem 1.** Given a neuron diagram with its corresponding equations \( M \), and an assignment to its variables \( V \). Consider the CP-logic theory \( T \), and a story \( b \), that we get when applying the translation discussed above. Then \( C \) is an actual cause of \( E \) in the diagram according to the definition given by Hall iff \( C \) is an actual cause of \( E \) in \( b \) and \( T \) according to definition 5.

**Proof.** See Appendix.

At first sight, definition 5 does not fit into the general framework we introduced earlier, as there is a quantifier over different branches instead of a single probability. However, we will now show that for a significant group of cases it actually suffices to consider just a single \( T' \), which can be described in terms of irrelevant and intrinsic laws.

Rather than looking at all of the reductions separately, we single out a minimal structure which contains the essence of our story. In general such a minimal structure need not be unique, as the story may contain elements none of which are necessary by themselves yet without all of them the essence is changed. To make this more precise, we have a look at the necessary elements.

**Definition 7.** A law \( r \) is necessary iff

- \( \forall d \in \text{Red}^{(C,E)}_b \colon r \in \text{Laws}_d \) and
- \( \forall d, e \in \text{Red}^{(C,E)}_b \colon r_d = r_e \).

We define \( \text{Nec}(b) \) as the set of all necessary laws.

In general it might be that there are two (or more) edges which are unnecessary by themselves, but at least one of them has to be present. Consider for example a case where \( C \) causes both \( A \) and \( B \), and each of those in return is sufficient to cause \( E \). Then neither the law \( A \) nor the law \( B \) is necessary, yet at least one of them has to be applied to get \( E \). In cases where this complication does not arise, we shall say that the story is simple.

**Definition 8.** A story \( b \) is said to be simple iff the following holds:

- \( \forall r \in \text{Laws}_b \colon \text{the head of } r \text{ contains at most two disjuncts} \)
- \( \forall d \in \text{Red}^{(C,E)}_b \) for all non-deterministic \( r \) in \( \text{Laws}_d \setminus \text{Nec}(b) \) : \( \exists e \in \text{Red}^{(C,E)}_d \) so that \( e = d \uparrow \text{ to the application of } r \) and \( r_d = r_e \).

As an example, note that the story in the previous paragraph is not simple. Neither the law \( A \) nor the law \( B \) is necessary. Now consider the \((C,E)\)-reduction where first \( B \) fails to cause \( C \), followed by \( A \) causing \( A \), which in turn causes \( E \). Then there is no \((C,E)\)-reduction that is identical to it up to and including the failure to cause \( B \), and yet afterwards fails to cause \( A \).

We are now in a position to formulate a theorem that will allow us to adjust Hall’s definition into our framework.

**Theorem 2.** If \( (\exists d \in \text{Red}^{(C,E)}_b) : P_d(-E|do(-C)) > 0 \) then \( P_{\text{do}(\neg C)}(-E) > 0 \). If \( b \) is simple, then the reverse implication holds as well.

**Proof.** See Appendix.

It is possible to add an additional criterion to turn this theorem into an equivalence that also holds for non-simple stories, however we would then likewise have to extend the description of our framework for causation. We choose not to do this, for two reasons. First of all, all stories that can be expressed by neuron diagrams, already satisfy the first requirement of simplicity. Second, all of the examples Hall discusses are simple, as are all of the classical examples discussed in the literature, such as Early and Late Preemption, Symmetric Overdetermination, Switches, etc.

As a result of this theorem, rather than having to look at all \((C,E)\)-reductions and calculate their associated probabilities, we need only find all the necessary laws and calculate a single probability. If the story \( b \) is simple, then this probability represents an extension of Hall’s definition, since they are equivalent if one ignores the value of the probability but for it being 0 or not. To obtain a workable definition of actual causation, we present a more constructive description of necessary laws. From now on we call the node resulting from the application of a law \( r \) in \( b \) \( \text{Node}_r^b \).

---

2For example, one could demand that also \( P_{\text{do}(\neg C)}(-E|do(-C)) \neq P_{\text{do}(\neg C)}(-E|do(C)) \).
Theorem 3. If \( b \) is simple, then a non-deterministic law \( r \) is necessary iff there is no \( (\mathcal{C}, E) \)-reduction passing through a sibling of \( \text{Node}_b \).

Proof. See Appendix.

With this result, we can finally formulate our version of Hall’s definition, which we will refer to as Hall07.

Hall07-Irrelevant. No law \( r \) of \( T \) is irrelevant.

Hall07-Intrinsic. A law \( r \) of \( T \) is intrinsic iff \( r \) happened in \( b \), and there is no branch \( d \) passing through a sibling of \( \text{Node}_b \) such that \( \{C, E\} \subseteq \text{Leaf}_d \subseteq \text{Leaf}_b \).

3.4 Hall 2004 Definitions

In (2004) Hall claims that it is impossible to account for the wide variety of examples in which we intuitively judge there to be actual causation by using a single, all-encompassing definition. Therefore Hall attempts to do so by defining two different concepts which both deserve to be called forms of causation but are nonetheless not co-extensive.

Dependence The first of these is counterfactual dependence given the story, or simply Dependence, as stated in Definition 2. It consists in the most straightforward way one can adapt the theory \( T \) so that it accommodates the information of the actual events: by fixing the disjuncts of all non-deterministic laws in correspondence to the observed story, giving \( T^\emptyset \). When we are dealing with a theory \( T \) in which the only non-deterministic laws are those representing exogenous variables – i.e., laws such as \( U : p \leftarrow \) – then clearly Hall’s notion of Dependence is equivalent to ours. In other cases, our notion provides a probabilistic extension of his. This definition of actual causation translates into the following conditions for irrelevance and intrinsicness.

Dependence-Irrelevant. No law \( r \) of \( T \) is irrelevant.

Dependence-Intrinsic. A law \( r \) of \( T \) is intrinsic iff \( r \) happened in \( b \).

Production The second concept is harder to grasp. It tries to express the idea that to cause something is to bring it about, or to produce it. Hall provides ample motivation for fleshing out the concept as he does, which we will not go into here. The precise definition of this concept – which can be found in the appendices – involves some technical elaborations, but for our discussion the following informal but intuitively clear definition suffices to understand it in the context of neuron diagrams: \( C \) is a producer of \( E \) iff there is a directed path of firing neurons in the diagram from \( C \) to \( E \).

In this diagram according to Hall iff \( C \) is a producer of \( E \) in \( b \) and \( T \) according to the CP-logic version stated here.

Proof. See Appendix.

Besides providing a probabilistic extension, the CP-logic version of production also offers a way to make sense of causation by omission. That is, just as with all of the definitions in our framework in fact, we can extend it to encompass negative literals such as \( \neg C \) to be causes as well.

4 Comparison

Table 1 presents a schematic overview of the four definitions discussed so far, as well as two new ones, that we have given appropriate names. The columns and rows give the criteria for a law \( r \) of \( T \) to be considered intrinsic, respectively irrelevant, in relation to a story \( b \), and an event \( E \). By \( r \leq_b E \), we denote that \( r \) happened in \( b \) before the literal \( E \) was caused.

| Irrelevant | Intrinsic |
|------------|-----------|
| \( \emptyset \) | Dependence Hall07 |
| \( \exists d : (d = b \text{ up to } E) \land r \geq_d E \) \( r \not_\leq_b E \lor r_\leq_b r \) | BV12 BV07 |
| Production | Production07 |

In order to illustrate the working of the definitions and to highlight their differences, we present an example:

Assassin decides to poison the meal of a victim, who subsequently Dies right before dessert. However, Murderer decided to murder the victim as well, so he poisoned the dessert. If Assassin had failed to do his job, then Backup would have done so all the same.

The causal laws that form the context of this story are give by the following theory:

\[
\begin{align*}
\text{Assassin} & : p \iff . \\
\text{Murderer} & : q \iff . \\
\text{Backup} & : r \iff \neg \text{Assassin}. \\
\text{Dies} & \iff \text{Assassin}. \\
\text{Dies} & \iff \text{Murderer}. \\
\text{Dies} & \iff \text{Backup}. \\
\end{align*}
\]

In this story, did Assassin cause Dies? We leave it to the reader to verify that in this case the left intrinsicness condition from the table applies to the first two non-deterministic laws, whereas the right one only applies to the first. The second irrelevance condition only applies to laws (17), whereas the third one also applies to laws (15) and (18). This results in the following probabilities representing the causal status of Assassin:

| Production | BV12 | Hall07 | Dependence |
|------------|------|-------|------------|
| 1 \( \leq r \) \( (1 - r) + (1 - q) \) 0 |

Different motivations can be provided for these answers:

- **Production**: Assassin brought about the death of the victim all by himself, hence he is the full cause.
• **BV12**: If *Assassin* hadn’t killed him, then that action would have not brought about victim’s death with a chance of \((1 - r)\). His action took away that option, hence he is the cause to that extent.

• **Hall07**: As far as *Assassin* could have known, there was a \((1 - r) \times (1 - q)\) chance that without his action victim would have survived. Hence he is the cause to that extent.

• **Dependence**: The victim would have died anyway, so *Assassin* is not a cause at all.

Rather than saying that only one of these answers is correct, we prefer to think of them as answering different questions, all of which have their use in some context or other. (Eg., to determine responsibility, understand *Assassin’s* state of mind, minimize the chance of murders, etc.) More generally, the definitions could be characterized by describing which events are allowed to happen in the counterfactual worlds they take into consideration to judge causation. Recall that we take an event to be the application of a law. We say that an event happens differently, if a different disjunct is chosen.

• **Production**: Only those events which led to \(E\), and not differently.

• **BV12**: Only those events which led to \(E\), and not differently, and also those events which were prevented from happening by these.

• **Hall07**: Any event can happen, as long as those events that were essential to lead to \(E\) do not happen differently.

• **Dependence**: Any event can happen, as long as those events that did actually happen do not happen differently.

## 5 Conclusion

In this paper we have used the formal language of CP-logic to formulate a general definition of actual causation, which we used to express four specific definitions: a proposal by Beckers and Vennekens, and three definitions based on the work of Hall. By moving from the deterministic context of neuron diagrams to the non-deterministic context of CP-logic, the latter definitions improve on the original in two ways: they can deal with a wider class of examples than the first, and they allows for a gradual judgment of actual causation in the form of a conditional probability. Further, comparison between the definitions is facilitated by presenting them as various ways of filling in two central concepts. On a whole our goal was to illustrate the flexibility of CP-logic in expressing different definitions, opening the path to other proposals beyond the ones here discussed.

## A Appendices

To facilitate the proof of the first theorem, we introduce the following lemma.

**Lemma 1.** Given a neuron diagram \(D\) with its corresponding equations \(M\), and an assignment to its variables \(V\). Consider the CP-logic theory \(T\), and a story \(b\), that we get when applying the translation discussed above. Then a neuron diagram \(R\) is a reduction of \(D\) in which both \(C\) and \(E\) occur \(iff\) its translation \(d\) – another branch of \(T\) – is a \((C,E)\)-reduction of \(b\). (As in definition 4.)

**Proof.** Assume we have a reduction \(R\) of a neuron diagram \(D\), and \(b\) is the story corresponding to \(D\). As \(R\) is simply a different assignment of the variables occurring in \(D\), brought about by the same equations that existed for \(D\), this reduction corresponds to another branch \(d\) of \(T\), in which \(C\) and \(E\) hold in its leaf. Moreover, \(R\) can be constructed starting from \(D\) by changing some of the exogenous variables, say \(U^*\), from their actual values to their default value, and then updating the endogenous variables in accordance with the deterministic equations. It being a reduction, this caused no new variables to take on their deviant value in comparison to \(D\). Let \(r\) be a law that occurs in \(d\).

If \(r\) is non-deterministic, it must be one of the laws representing an exogenous variable \(V\), i.e., a law with an empty body, and hence it was also applied in \(b\). \(R\) being a reduction, either \(V\) has the same value in \(R\) as in the original diagram, or it has its default value. In the former case, this means that \(r_d = r_b\), in the latter case \(r_d = 0\), both of which satisfy the requirement for \(d\) being a reduction.

If \(r\) is deterministic, the precondition for \(r\) has to be fulfilled in \(d\), causing some variable \(V\) to take on its deviant value. The same must hold true of the precondition for the equation for \(V\), and thus \(V\) takes on its deviant value in \(R\) as well, implying it did so in \(D\) too. Therefore there must have been some law applied in \(b\) that made \(V\) take on its deviant value as well. From this it follows that \(d\) is a \((C,E)\)-reduction of \(b\).

Now assume we have a theory \(T\) and a story \(b\) that form the translation of a neuron diagram \(D\), such that \(C\) and \(E\) hold in \(b\), and that \(d\) is a \((C,E)\)-reduction of \(b\). As the leaf of \(d\) contains an assignment to all of the variables that satisfies the equations of \(M\), there is a neuron diagram \(R\) that corresponds to \(d\). We can easily go over all the previous steps in the other direction, to conclude that \(R\) is a reduction of \(D\) in which \(C\) and \(E\) are true.

**Theorem 1.** Given a neuron diagram with its corresponding equations \(M\), and an assignment to its variables \(V\). Consider the CP-logic theory \(T\), and a story \(b\), that we get when applying the translation discussed above. Then \(C\) is an actual cause of \(E\) in the diagram according to the definition given by Hall iff \(C\) is an actual cause of \(E\) in \(b\) and \(T\) according to definition 5.

**Proof.** We start with the implication from left to right. Assume we have a neuron diagram \(D\), in which both \(C\) and \(E\) fire. This translates into a theory \(T\) and a story \(b\), for which \(C\) and \(E\) hold in its leaf. Further, assume there is a reduction \(R\) of this diagram, in which both \(C\) and \(E\) continue to hold, and in this reduction, if \(C = 0\); then \(E = 0\). By the above lemma, this translates into a \((C,E)\)-reduction of \(b\), say \(d\).

In \(R\), if \(C = 0\); then \(E = 0\). The conditional \(C = 0\) is interpreted as a counterfactual locution, and corresponds to \(do(\neg C)\). As there are no non-deterministic laws with non-empty preconditions, \(T^d\) is simply the deterministic
theory that determines the same assignment as \( R \), meaning \( P_{P_d}(\neg E|do(\neg C)) = 1 \), which concludes this part of the proof.

Now assume we have a theory \( T \) and a story \( b \) that form the translation of a neuron diagram \( D \), such that \( C \) and \( E \) hold in \( b \), and that \( d \) is a \((C,E)\)-reduction of \( b \) for which the given inequality holds. By the above lemma, the translation of \( d \), say \( R \), is reduction of \( D \) in which \( C \) and \( E \) occur. As mentioned in the previous paragraph, \( T^d \) simply corresponds to an assignment of values to the variables occurring in \( D \) that follows its equations. Since \( R \) describes this same assignment, in \( R \) too if \( C = 0 \); then \( E = 0 \). This concludes the proof.

**Theorem 2.** If \( (\exists d \in \text{Red}_{b}(C,E) : P_{P_d}(\neg E|do(\neg C)) > 0) \) then \( P_{P_{\text{Nec}}}(\neg E|do(\neg C)) > 0 \). If \( b \) is simple, then the reverse implication holds as well.

*Proof.* We start with proving the first implication. Assume we have \( a,d \in \text{Red}_{b}(C,E) \) such that \( P_{P_d}(\neg E|do(\neg C)) > 0 \). This implies that there is at least one branch \( e \) of a probability tree of \( T^d|do(\neg C) \) for which \( \neg E \) holds in its leaf. We prove by induction on the length of \( e \) that this implies the existence of a similar branch \( e' \) of a probability tree of \( T_{\text{Nec}}(b)|do(\neg C) \) for which \( \neg E \) holds in its leaf, which is what is required to establish the theorem.

Base case: if \( e \) consists of a single node – i.e., the root node where all atoms are false – then this means that no laws of \( T^d|do(\neg C) \) can be applied. Since the bodies of the laws in \( T_{\text{Nec}}(b)|do(\neg C) \) are identical to those of the laws in \( T^d|do(\neg C) \), we simply have \( e' = e \).

Induction case: Assume we have a sub-branch \( e_n \) of \( e \) with length \( n > 1 \), starting from the root node, and that we also have a structurally identical sub-branch \( e'_n \). By it being structurally identical we mean that they are identical except for the fact that they may have different probabilities along the edges.

If \( e_n = e \), then no more laws can be applied in the final node of \( e_n \). This must then hold for the final node of \( e'_n \) as well, so we are finished. Otherwise, we know that there is a sub-branch \( e_{n+1} \) which extends \( e_n \) along \( e \) with a node \( O \). Assume that the law which was applied to get to \( O \) is \( r \).

If \( r \) is deterministic, then \( r \) occurs in \( T^d|do(\neg C) \) exactly as it does in \( T_{\text{Nec}}(b)|do(\neg C) \). Since both branches are structurally identical, \( e'_n \) can be extended in the exact same manner as \( e_n \), so there has to be a probability tree of \( T_{\text{Nec}}(b)|do(\neg C) \) in which there is a sub-branch \( e'_{n+1} \) with the desired properties. So assume \( r \) is non-deterministic.

First assume \( r \not\in \text{Laws}_{f} \). This implies that \( r \not\in \text{Nec}(b) \). So as in the deterministic case, \( r \) occurs in \( T^d|do(\neg C) \) exactly as it does in \( T_{\text{Nec}}(b)|do(\neg C) \), and the branch can be extended in the same manner.

Now assume \( r \in \text{Laws}_{d} \). If also \( r \in \text{Nec}(b) \), we know that \( t_d = t_b = t_{\text{Nec}} \), and hence the previous argument holds. Remains the possibility that \( r \not\in \text{Nec}(b) \). As in the deterministic case, because \( r \) can be applied in the final node of \( e'_n \) there has to be a probability tree of \( T_{\text{Nec}}(b)|do(\neg C) \) with a sub-branch like \( e'_{n+1} \) where \( r \) is applied next.

Assume \( r_d = A \). Since \( A \) was the outcome of \( r \) in \( d \), the law \( r \) as it appears in \( T \) – and also in \( T_{\text{Nec}}(b)|do(\neg C) \) – contains \( A \) in its head with some probability attached to it. Therefore the final node of \( e'_{n+1} \) in the said probability tree has one child-node which contains \( A \), extending \( e'_{n+1} \) into a sub-branch \( e'_{n+1} \) with the desired properties. This concludes this part of the proof.

Now we prove that if \( b \) is simple, the reverse implication holds as well. Assume \( P_{P_{\text{Nec}}}(\neg E|do(\neg C)) > 0 \). This implies that there is at least one branch \( e \) of a probability tree of \( T_{\text{Nec}}(b)|do(\neg C) \) for which \( \neg E \) holds in its leaf. We can repeat the first steps of the previous implication, so that we again arrive at a law \( r \) which was applied to get to a node \( O \).

The branch \( e' \) we are considering occurs in a probability tree of a \((C,E)\)-reduction, say \( f \). First assume \( r \in \text{Nec}(b) \). By definition, this implies that also \( r \in \text{Laws}_{f} \) and \( r_{\text{Nec}} = r \), and we can apply the reasoning from above. Likewise as above, we can apply this reasoning to all other cases, except the one where \( r \not\in \text{Nec}(b) \), \( r \) is non-deterministic, and \( r \in \text{Laws}_{f} \). Assume the law \( r \) has effect \( A \) in the branch \( e \) we are considering. If \( r_f = A \), then we are back to our familiar situation, so therefore assume \( r_f = B \), and \( A \not= B \).

Since \( b \) is simple, \( A \) and \( B \) are the only two possible effects of \( r \). Further, remark that \( r \in \text{Laws}_{b} \setminus \text{Nec}(b) \). This implies the existence of a \((C,E)\)-reduction \( g \) that is identical to \( f \) up to the application of \( r \), but such that \( r \not= r_f \), and thus \( r_g = A = r_e \) meaning there is a branch in a probability tree of \( g \) that is structurally identical to \( e \) up to \( O \). This concludes the proof of the theorem.

**Theorem 3.** If \( b \) is simple, then a non-deterministic law \( r \) is necessary iif there is no \((C,E)\)-reduction passing through a sibling of \( \text{Node}_{b} \).

*Proof.* Say the unique sibling of \( \text{Node}_{b} \) is \( M \). We start with the implication from left to right, so we assume \( r \) is necessary. Assume \( r_b = A \), then there is no \( d \in \text{Red}_{b}(C,E) \) for which \( t_d \not= A \), hence there is no \((C,E)\)-reduction which passes through \( M \).

Remains the implication from right to left. Assume we have a law \( r \) such that there is no \((C,E)\)-reduction passing through a sibling of \( \text{Node}_{b} \). We proceed with a reductio ad absurdum, so we assume \( r \) is not necessary.

Clearly \( b \) is a \((C,E)\)-reduction of itself, and also \( r \in \text{Laws}_{b} \setminus \text{Nec}(b) \). Hence, by \( b \)’s simplicity, there is a \((C,E)\)-reduction \( e \) which is identical to \( b \) up to the application of \( r \), but for which \( r_e \not= r_b \). Thus \( e \) passes through the sibling of \( \text{Node}_{b} \), contradicting the assumption that \( r \) is necessary. This concludes the proof.

**Theorem 4.** Given a neuron diagram with its corresponding equations \( M \), and an assignment to its variables \( V \). Consider the CP-logic theory \( T \), and a story \( b \), that we get when applying the translation discussed earlier. \( C \) is a producer of \( E \) in the diagram according to Hall if \( C \) is a producer of \( E \) in \( b \) and \( T \) according to the CP-logic version stated here.

*Proof.* First we need to explain some terminology that Hall uses. A structure is a temporal sequence of sets of events,
which unfold according to the equations of some neuron diagram. A branch, or a sub-branch, would be the corresponding concept in CP-logic.

Two structures are said to match intrinsically when they are represented in an identical manner. The reason why Hall uses this term, is because even though we use the same variable for an event occurring in different circumstances, strictly speaking they are not the same. This is mainly an ontological issue, which need not detain us for our present purposes.

A set of events $S$ is said to be sufficient for another event $E$, if the fact that $E$ occurs follows from the causal laws, together with the premise that $S$ occurs at some time $t$, and no other events occur at this time. A set is minimally sufficient if it is sufficient, and no proper subset is. To understand this, note that the ambiguity of the relation between an event and the value of a variable that we noted earlier, resurfaces here. In the context of neuron diagrams, events are temporal, and the value of a variable that we noted earlier, resurfaces. Interpretation, it is natural to translate Hall’s notion of an event “the neuron fires” to “the neuron has fired”. Given this interpretation, it is natural to translate Hall’s notion of an event into CP-logic as the application of a law, making a variable true, as we have done.

A further detail to be cleared out, is that in the context of neuron diagrams there can be simultaneous events, since multiple neurons can fire at the same time. In CP-logic, in each node only one law is allowed to happen, hence this translates to two consecutive edges in a branch. Therefore it is not the case that each node-edge pair in a branch corresponds to a separate time-point, but rather sets of consecutive pairs – with variable size – do. Given such a set, then for each variable that was the result of the application of a law belonging to it, it holds that its corresponding event occurs at the next time-point, corresponding to the next set of nodes further down the branch. All the variables occurring in the bodies of the laws in this set, represent events that occur during this time-point.

Now we can state the precise definition of production as it occurs in (Hall 2004, p.25).

We begin as before, by supposing that $E$ occurs at $t'$, and that $t$ is an earlier time such that at each time between $t$ and $t'$, there is a unique minimally sufficient set for $E$. But now we add the requirement that whenever $t_0$ and $t_1$ are two such times $(t_0 < t_1)$ and $S_0$ and $S_1$ the corresponding minimally sufficient sets, then

- for each element of $S_1$, there is at $t_0$ a unique minimally sufficient set; and
- the union of these minimally sufficient sets is $S_0$.

Given some event $E$ occurring at time $t'$ and given some earlier time $t$, we will say that $E$ has a pure causal history back to time $t$ just in case there is, at every time between $t$ and $t'$, a unique minimally sufficient set for $E$, and the collection of these sets meets the two foregoing constraints. We will call the structure consisting of the members of these sets the “pure causal history” of $E$, back to time $t$. We will say that $C$ is a proximate cause of $E$ just in case $C$ and $E$ belong to some structure of events $S$ for which there is at least one non-ontologically possible structure $S'$ such that (i) $S'$ intrinsically matches $S$; and (ii) $S'$ consists of an $E$-duplicate, together with a pure causal history of this $E$-duplicate back to some earlier time. (In easy cases, $S$ will itself be the needed duplicate structure.) Production, finally, is defined as the ancestral [i.e., the transitive closure] of proximate causation.

We will start with the implication from left to right. So assume we have a neuron diagram $D$, in which $C$ is a producer of $E$. Say $T$ is the CP-logic theory that is the translation of the equations of the diagram, and $b$ is the branch representing the story. We already know that $C$ and $E$ hold in the leaf of $b$. We need to proof that $P_f(\neg E | do(\neg C)) > 0$. The theory $T'$ only contains deterministic laws, and no disjunctions, hence all its laws are of the form: $V \leftarrow A \land A' \land ... \land B \land \neg B'$, where the number of positive literals in the conjunction is at least one. Therefore any probability tree for $T'$ consists out of only one branch, determining a unique assignment for all the variables. Further, even though the theory $T$ may contain several laws in which a variable occurs in the head, because of our irrelevance criterion $T'$ contains exactly one law for every variable that is true. So for every true variable in this assignment, there is a unique chain of laws – neglecting the order – which need to happen to make this variable true. For any such variable $V$, we will say that it depends on all of the variables occurring positively in the body of a law in this chain. Clearly, if any true variable changes its value in this assignment, then all variables which depend on it become false.

As a first case, assume $C$ is a proximate cause of $E$. We start by assuming that circumstances are nice, meaning that $D$ contains itself a structure $S$ which is a pure causal history of $E$. This means that in the actual story $b$, $C$ is part of a unique minimally sufficient set for $E$. From this it follows that in $T'$, $C$ figures positively in one of the laws on which $E$ depends. Hence, if we apply $do(\neg C)$, then $E$ will no longer hold.

Now assume that there is a structure $S$ occurring in $D$, such that there exists another diagram, say $D'$, in which this structure occurs as well, and forms a pure causal history of $E$. This diagram corresponds to a branch of $T'$, say $d$. That means that in $T'_d$, the theory $T'$ constructed out of $d$ is $C$ occurs positively in the unique chain of laws which can make $E$ true. But as all events in $S$ also occur in $D$, at the same moments as they do in $D'$, that means that $C$ must also occur positively in the unique chain of laws for $E$ in the theory $T'_d$. Hence, $E$ depends on $C$ in the theory $T'_d$ as well.

Now look at the more general case, in which $C$ occurs in a chain of proximate causes, that leads up to $E$. I.e., in $D$, $C$ is the proximate cause of some variable $V_1$, which in turn is the proximate cause of some variable $V_2$, and so on until we get to $E$. We know from the previous discussion, that this implies in $T'$ that $do(\neg C)$ then $\neg V_1$, and $do(\neg V_1)$ then $\neg V_2$, and so on. Given what we know about $T'$, it directly follows...
that when we apply \( \text{do}(-C) \), then \(-E\). This concludes this part of the proof.

We continue with the implication from right to left. So assume that we are given again a neuron diagram and a corresponding story \( b \), and that we know \( \Pr((-E|O\text{do}(-C))) > 0 \). From our earlier analysis of \( T' \), we know that this means that \( C \) occurs positively in the unique chain of laws that can make \( E \) true according to \( T' \). From this chain of laws, we start from the one causing \( E \) and from there pick out a series that gets us to a law where \( C \) occurs positively in the body. More concretely, we take a series of the form: \( E \leftarrow \ldots A \land \ldots, A \leftarrow \ldots D \land \ldots \), and so on until we get at a law \( Z \leftarrow \ldots C \land \ldots \).

By definition of production, it suffices to prove that in this chain, each of the variables in the body is a proximate cause of the variable in the head.

Take such a law \( V \leftarrow \ldots W \land \ldots \). At the time that this law is applied, \( W \) clearly is a member of a sufficient set of events for \( V \), which occurs at the next time point. Say \( S_0 \) is the set of all events that occur together with \( W \) that figure in the body of this law, and \( S_1 \) is the set \( \{V\} \) that occurs at the next time-point, then the structure consisting precisely of \( S_0 \) and \( S_1 \) and nothing else forms a pure causal history of \( V \) containing \( W \). The same reasoning applies to all laws of the chain. This concludes the proof.

References

Beckers, S., and Vennekens, J. 2012. Counterfactual dependency and actual causation in cp-logic and structural models: a comparison. In Kersting, K., and Toussaint, M., eds., Proceedings of the Sixth STAIRS, volume 241 of Frontiers in Artificial Intelligence and Applications, 35–46.

Hall, N., and Paul, L. A. 2003. Causation and Preemption. Oxford University Press.

Hall, N. 2004. Two concepts of causation. In Causation and Counterfactuals.

Hall, N. 2007. Structural equations and causation. Philosophical Studies 132(1):109–136.

Halpern, J. Y., and Hitchcock, C. Graded causation and defaults. British Journal for the Philosophy of Science forthcoming. http://arxiv.org/pdf/1309.1226v1.pdf.

Halpern, J., and Pearl, J. 2005a. Causes and explanations: A structural-model approach. part I: Causes. The British Journal for the Philosophy of Science 56(4):843–87.

Halpern, J., and Pearl, J. 2005b. Causes and explanations: A structural-model approach. part II: Explanations. The British Journal for the Philosophy of Science 56(4).

Hitchcock, C. 2007. Prevention, preemption, and the principle of sufficient reason. Philosophical review 116(4):495–532.

Hitchcock, C. 2009. Structural equations and causation: six counterexamples. Philosophical Studies 144:391–401.

Lewis, D. 1973. Causation. J. of Philosophy 70:113–126.

Pearl, J. 2000. Causality: Models, Reasoning, and Inference. New York, NY, USA: Cambridge University Press.

Shafer, G. 1996. The Art of Causal Conjecture. Artificial Management. MIT Press.

Vennekens, J.; Denecker, M.; and Bruynooghe, M. 2009. CP-logic: A language of probabilistic causal laws and its relation to logic programming. Theory and Practice of Logic Programming 9:245–308.

Vennekens, J.; Denecker, M.; and Bruynooghe, M. 2010. Embracing events in causal modelling: Interventions and counterfactuals in CP-logic. In JELIA, 313–325.

Vennekens, J. 2011. Actual causation in cp-logic. Theory and Practice of Logic Programming 11:647–662.