Causation in Terms of Production[*]

Holger Andreas & Mario Günther

Abstract

In this paper, we analyse actual causation in terms of *production*. The latter concept is made precise by a strengthened Ramsey Test semantics of conditionals: $A \gg C$ iff, after suspending judgement about A and C, C is believed in the course of assuming A. This test allows us to (epistemically) verify or falsify that an event brings about another event. Complementing the concept of production by a weak condition of difference-making gives rise to a full-fledged analysis of causation.

Contents

[*] The authors contributed equally to this paper.

1 Introduction

In this paper, we develop a logical analysis of actual causation in terms of *production*. Inspired by Hall (2004, 2007), we regard production rather than counterfactual dependence to be more central to causation. We analyse the concept of production by a strengthened Ramsey Test conditional. Based on the Ramsey Test conditional, we define a concept of causation that solves the problems of overdetermination, preemption, and switches.

The evaluation recipe of our strengthened Ramsey Test conditional can be expressed as follows:

> First, suspend judgement about the antecedent and the consequent. Second, add the antecedent (hypothetically) to your stock of explicit beliefs. Finally, consider whether or not the consequent is entailed by your explicit beliefs.

In brief, $A \gg C$ iff, after suspending judgement about A and C, C is believed in the course of assuming A. We suggest that such a test allows us to (epistemically) verify or falsify that an event brings about a certain other event, and thus qualifies as a candidate cause of the latter event. Hence, as a preliminary starting point, we define:

$$C \text{ is a cause of } E \text{ iff } C \text{ and } E \text{ occur, and } C \gg E. \qquad \text{(Def C)}$$

The logical foundations of the belief changes that define the conditional $\gg$ are explicated using AGM-style belief revision theory, as founded by Alchourrón et al. (1985) and fleshed out by Gärdenfors (1988).

We aim to set forth the first analysis of causation in terms of production that is formally as rigorous as the counterfactual accounts of causation by Lewis (1973) and Halpern and Pearl (2005). As pointed out by Paul and Hall (2013), these counterfactual accounts face persistent problems, especially the combination of overdetermination, preemption, and switches. For our analysis the problem of overdetermination does not even arise. Early and late preemption turn out to be the same problem, which is solved by imposing constraints on the inferential relations underlying the conditional $\gg$. Halpern and Pearl (2005) solve the problems of overdetermination and preemption in a formally rigorous way as well, but at the price of rather complex set-theoretic constructions that are difficult to supplement with an intuitive motivation. For lack of space, we do not discuss causal models, and will compare our analysis in greater detail with the one of Halpern and Pearl (2005) elsewhere.

The plan of our investigation is straightforward: we work upward from belief changes via the strengthened Ramsey Test to an analysis of causation. In Section 2, we explain the basic ideas of AGM-style belief revision theory and the Ramsey Test before moving on to our strengthened Ramsey Test. On the basis of the strengthened Ramsey Test conditional $\gg$, Section 3 develops an epistemic approach to causation. In a stepwise fashion, we deal with the problems of overdetermination, early and late preemption, and switches. Section 4 concludes the paper.

2 Belief Changes and the Ramsey Test

2.1 Belief Changes: Basic Ideas

AGM-style belief revision theory provides a precise semantics of belief changes for the Ramsey Test. Let us therefore make ourselves familiar with the basic ideas of belief revision. Suppose K is a set of formulas that represent the beliefs of an agent, while A is a formula that represents a single belief. In the AGM framework, one distinguishes three types of belief change of a belief set K by a formula A:

(1) Expansions $K + A$

(2) Revisions $K * A$

(3) Contractions $K \div A$.

An expansion of K by A consists in the addition of a new belief A to the belief set K. This operation is not constrained by any considerations as to whether the new epistemic input A is consistent with the set K of present beliefs. Hence, none of the present beliefs is retracted by an expansion. $K + A$ designates the expanded belief set.

A revision of K by A, by contrast, can be described as the *consistent integration* of a new epistemic input A into a belief system K. If A is consistent with K, it holds that $K + A = K * A$, i. e. the revision by A is equivalent to the expansion by A. If, however, A is not consistent with K, some of the present beliefs are to be retracted, as a consequence of adopting the new epistemic input. $K * A$ designates the revised system of beliefs.

A contraction of K by A, finally, consists in retracting a certain formula A from the presently accepted system of beliefs. This operation will be used to define the *suspension of judgement about A* in our strengthened version of the Ramsey Test. $K \div A$ designates the belief set after the retraction of A.

In some contexts, it is helpful to distinguish between the belief system K and the epistemic state S that underlies it. Henceforth, we shall make this distinction, and write $K(S)$ for the belief system K of the epistemic state S.

Belief changes can be defined in various ways. A large number of different belief revision schemes have been developed in the spirit of the original AGM theory. We shall assume that epistemic states are represented by *belief bases*. In symbols, $S = H$. A belief base H is a set of formulas that represent the explicit beliefs of an agent. Belief base revision schemes are guided by the idea that the inferential closure of a belief base H gives us the belief set K of H:

$$K(H) =_{df} Inf(H).$$

K contains all beliefs of the epistemic state H, i. e. the explicit beliefs and those beliefs that the agent is committed to accept because they are inferable from the explicit beliefs. Inf is an inferential closure operation that contains classical logic (see Gärdenfors (1988) and Hansson (1992)). In this paper, we follow suit in making the conservative assumption that Inf is given by the consequence operation of classical logic Cn. Hence, the belief set $K(H)$ is defined by $Cn(H)$.

The definition of an expansion is straightforward for belief bases:

$$K(H) + A =_{df} K(H + A)$$

where $H + A$ stands for adding the new epistemic input to the belief base H.

Note that we can define revisions in terms of contractions and expansions:

$$K(S) * A = (K(S) \div \neg A) + A. \qquad \text{(Levi identity)}$$

Once we have retracted $\neg A$, we obtain a belief set $K(S')$ that is consistent with A. Hence, we have $K(S') * A = K(S') + A$.

In the following analysis of causation, we assume that the belief bases have exactly two levels of epistemic priority: the upper level, containing the generalisations and implications that describe the dynamics of events, and the lower level, which contains our beliefs about atomic facts. These levels of epistemic priority affect the determination of belief changes: when we retract a belief A, we first retract atomic beliefs that imply A (in the context of the generalisations and implications). If necessary, we also retract generalisations and implications, but only if the retraction of A cannot be achieved by retractions of beliefs with lower epistemic priority. For the considerations to follow, it may be helpful to have a graphical representation of such a prioritised belief base in mind:

$G \cup I$
F

G stands for the set of generalisations, I for the set of implications, and F contains the beliefs about the atomic facts. For some causal scenarios, it is necessary to distinguish between strict and *ceteris paribus* generalisations. Then, we need to have at least two levels of epistemic priority for the members of $G \cup I$. The canonical scenarios of preemption, overdetermination, conjunctive causes, and switches can be represented using just one level of generalisations and implications. We shall define contractions and revisions for a prioritised belief base in the next section. For an intuitive understanding of our analysis, however, the informal explanations of belief changes in this section suffice.

2.2 Defining Belief Changes

In view of the Levi identity, we can focus on belief base contractions and thereby define a belief base revision scheme. Recall that the contraction of $K(S)$ by A yields a belief set $K(S')$ that does not contain A. A contraction by A can be defined using the notion of a *remainder set* $H \perp A$. The remainder set $H \perp A$ contains all maximal subsets of H that do not entail A. In formal terms:

Definition 1. $H \bot A$ (Hansson, 1999, p. 12)
Let H be a set of formulas and A a formula. $H' \in H \bot A$ iff

(1) $H' \subseteq H$

(2) $A \notin Cn(H')$

(3) there is no H'' such that $H' \subset H'' \subseteq H$ and $A \notin Cn(H'')$.

The next step is to define the contraction of a belief base H by A:

$$H \div_{\sigma} A =_{df} \sigma H \bot A. \qquad \text{(MC Contraction)}$$

σ stands for a *selection function* that picks out a specific element of the remainder set $H \bot A$. The contraction operation defined by such a selection function is also referred to as *maxichoice contraction*. Such a contraction allows for a maximally conservative way of retracting a belief, in the sense that as many as possible of the present beliefs are retained. We will see this behaviour of maxichoice contractions at work when we analyse a combination of a conjunctive causal scenario with overdetermination in Section 3.5.

Some beliefs are more firmly established than others. To respect such differences in defining belief contractions, we impose a constraint on the selection function, using levels of epistemic priority. Let

$$\mathbf{H} = \langle H_1, \ldots, H_n \rangle$$

be a prioritised belief base. That is, $H_1, \ldots, H_n$ are sets of formulas that represent explicit beliefs, and the indices represent an epistemic ranking of the beliefs. H_1 is the set of the most firmly established beliefs, the beliefs in H_2 have secondary priority, etc. We say that a selection function σ is epistemically superior to another selection function σ' iff there is an epistemic level i such that σ selects strictly more sentences of level i than σ', while being on a par with σ' as regards the levels $j < i$. To be more precise:

Definition 2. $\sigma < \sigma'$
Let $\mathbf{H} = \langle H_1, \ldots, H_n \rangle$ be a prioritised belief base. $H = H_1 \cup \ldots \cup H_n$. Let σ and σ' be two selection functions for the remainder set $H \bot A$. $\sigma < \sigma'$ iff

(1) there is i $(1 \leq i \leq n)$ such that $(\sigma' H \bot A) \cap H_i \subset (\sigma H \bot A) \cap H_i$, and

(2) for all j $(1 \leq j < i)$, $(\sigma' H \bot A) \cap H_j = (\sigma H \bot A) \cap H_j$.

Definition 3. Epistemically optimal selection function
We say that a selection function σ for $H \perp A$ is epistemically optimal – with respect to $\mathbf{H}$ – iff there is no selection function σ' for $H \perp A$ such that $\sigma' < \sigma$.

The basic idea underlying these two definitions is very simple: when retracting a belief A, we should retain the more firmly established beliefs, while the less firmly established beliefs may be more readily given up. More firmly established beliefs are to be given up only if there is no other way to retract the belief A. Note that there may be more than one epistemically optimal selection function for a given contraction problem. The requirement that the selection function be epistemically optimal merely constrains this function, but does not define it.

Now we are in a position to define maxichoice contractions of a prioritised belief base $\mathbf{H}$:

$$\mathbf{H} \div_{\sigma} A =_{df} \sigma H \perp A \qquad \text{(Def } \div_{\sigma}\text{)}$$

where $H = H_1 \cup \ldots \cup H_n$ and σ is epistemically optimal with respect to $\mathbf{H}$. Note that the result of contracting a prioritised belief base $\mathbf{H}$ is a non-prioritised belief base H. This is not very elegant, but greatly simplifies our analysis of causation.

The definition of an expansion of a non-prioritised belief base is straightforward:

$$H + A =_{df} \; H \cup \{A\}. \qquad \text{(Def +)}$$

Using the Levi identity, we obtain the definition of a revision of a prioritised belief base:

$$\mathbf{H} *_{\sigma} A =_{df} (\sigma H \perp \neg A) + A \qquad \text{(Def } *_{\sigma}\text{)}$$

where σ is epistemically optimal with respect to $\mathbf{H}$. In making the relativisation to a selection function notationally explicit, we deviate from standard AGM notations. This deviation, however, will be needed in what follows. Depending on the context, we shall continue to use the non-relativised notation as well. We have now all definitions at hand to introduce the strengthened Ramsey Test conditional underlying our analysis of causation.

2.3 The Ramsey Test

Ramsey (1929/1990) devised an epistemic evaluation recipe for conditionals that is nowadays known as the Ramsey Test. Its core idea has been pointedly expressed by Stalnaker (1968, p. 102):

> First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true.

It was then Gärdenfors (1978) who translated this test into the language of belief changes and who insisted more forcefully than Stalnaker on an epistemic understanding of conditionals. Using the AGM framework, he was able to define a semantics of conditionals explicitly in terms of belief changes:

$$A > C \in K(S) \text{ iff } C \in K(S) * A \tag{RT}$$

where > designates the conditional connective. Recall that $K(S) * A$ designates the revision of the beliefs of an epistemic state S with the formula A. So the Ramsey Test defines that a conditional $A > C$ is to be accepted in a belief system $K(S)$ iff the consequent C is in the revision of $K(S)$ by the antecedent A. Unlike Gärdenfors (1978), we require that A and C be non-conditional formulas.[1]

2.4 Strengthening the Ramsey Test

We define a conditional $\gg$ with the following intuitive meaning: $A \gg_\sigma C$ iff, after suspending any beliefs in $K(S)$ as to whether A and C are true or false (using σ), we can infer C from A, in the context of the remaining beliefs. In more formal terms:

Definition 4. Belief function $B(A)$
Let T be some arbitrary classical tautology and A a formula.

$$B(A) = \begin{cases} A & \text{if } A \in K(S) \\ \neg A & \text{if } \neg A \in K(S) \\ \neg T & \text{otherwise.} \end{cases}$$

[1] Gärdenfors (1986) has proven a triviality theorem concerning the Ramsey Test, after a conditional logic was developed on the basis of this test in Gärdenfors (1978). Recently, however, there have been various, apparently successful attempts at defending the Ramsey Test in light of this result (see, e.g., Bradley (2007)). We show that our strengthened Ramsey Test does not imply triviality in Andreas and Günther (2018). The following section draws on this paper, where the conditional $\gg$ has been defined for the first time.

$$A \gg_\sigma C \in K(S) \text{ iff } C \in (K(S) \div_\sigma B(A) \vee B(C)) + A. \qquad (\mathrm{SRT}_\sigma)$$

Equivalently,

$$A \gg_\sigma C \in K(S) \text{ iff } (K(S) \div_\sigma B(A) \vee B(C)), A \vdash C.$$

where $\vdash$ designates the relation of provability in classical logic. The first step of (SRT_σ) consists in an *agnostic move* that lets us suspend judgement about the antecedent and the consequent. Then, we check whether or not we can infer the consequent C from the antecedent A, in the context of the remaining beliefs of the epistemic state. If so, $A \gg_\sigma C \in K(S)$. Otherwise, $A \gg_\sigma C \notin K(S)$.

The conditional $\gg_\sigma$ is obviously relative to a selection function σ. To obtain our intended analysis of causation, we eliminate the relativisation by means of existential quantification:

$$A \gg C \in K(S) \text{ iff there is a } \sigma \text{ s.t. } C \in (K(S) \div_\sigma B(A) \vee B(C)) + A. \quad (\mathrm{SRT})$$

If the epistemic state S is furnished with some epistemic ordering of beliefs, σ must be epistemically optimal with respect to this ordering. This constraint is contained in our definition of $\div_\sigma$ for prioritised belief bases. Henceforth, we assume that there is a prioritised belief base $\mathbf{H}$ such that $S = \mathbf{H}$.

(SRT) is the formulation of our strengthened Ramsey Test that we eventually use in the present analysis of causation. For many causal scenarios, however, $\gg$ and $\gg_\sigma$ are equivalent since the epistemic constraint on σ suffices to define a selection function uniquely. In this case, there is only one epistemically optimal selection function. It is only certain specific causal scenarios that require the additional flexibility of $\gg$. We shall explain such a scenario in Section 3.5. We leave the selection function implicit in the applications of (SRT) whenever this function is uniquely determined by the requirement of being epistemically optimal.

3 Causation

3.1 Actual Causation

We are aiming at an analysis of actual causation between events. Actual causation concerns the question of whether or not a particular occurrent event causes another occurrent event. It is related to, but different from causation at the type-level, which concerns causal relations between repeatable events. Unless otherwise stated, upper case Latin letters stand for propositions saying that a certain event occurs. By abuse of notation, we shall also speak of the event C, and thereby refer to the event whose occurrence is claimed by the proposition C.

We express the requirement that the causal relata be occurring events as follows:

$$C, E \in K(S). \tag{C1}$$

$K(S)$ designates the belief set of the epistemic state S. Our analysis is thus relative to an epistemic state, in a similar way as the analysis of Spohn (2006) is, and perhaps the one of Hume (1739/1978).[2]

3.2 Production

Recall that our analysis of causation is driven by the idea that a cause brings about, or produces, its effect. Consequently, we advance the strengthened Ramsey Test as an inferential means to verify or falsify that a certain event brings about a certain other event. Hence, for C to be a cause of E it must hold that

$$C \gg E \in K(S).$$

This condition allows us to capture a large range of causal relations. It is somewhat too liberal, however. Sometimes, we are not only able to infer the effect from the cause, but also the other way around. For example, thunder caused by lightning seems to be unique in the sense that it is different from thunder caused by blasts or supersonic aircraft. (If it is not, then the better for our analysis.) Hence, thunder strongly conditionally implies – in the sense of (SRT) – lightning. But it seems counterintuitive and wrong to view thunder as a cause of lightning. For thunder does not produce lightning.

The idea of production seems to imply a temporal asymmetry between the producing event and the effect: the cause must precede its effect. Hence,

$$t(C) < t(E) \tag{C2}$$

[2] In the *Treatise*, Hume (1739/1978, p. 170, our emphasis) defines: “A cause is an object precedent and contiguous to another, and so united with it, that the idea of the one *determines the mind to form the idea of the other*”. In more modern terms, if an event C precedes another E (and is contiguous), Hume calls C a cause of E iff C is an epistemic reason for E. Observe that the notion of causation is relative to a ‘mind’.

where $t(C)$ is a function that yields the time at which the event C occurs. (C2) expresses an old Humean dictum on causation, which is also central to Spohn's ranking-theoretic analysis of causation. We thus take the temporal order of events as not relying on causal relations. Once (C2) is in place, our analysis rules out that the thunder is a cause of the lightning, as intended. If A, B, or A and B are temporally extended events, we take $t(A) < t(B)$ to mean that A comes into being before B, while A and B may well overlap.

3.3 Joint Effects

Joint effects of a common cause can pose a problem for an inferential approach to causation. Take the following neuron diagram from Paul and Hall (2013, p. 71):

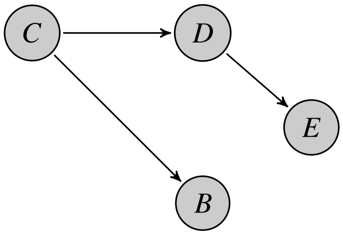

Figure 1

C fires and, thereby, sends signals to D and B so that D and B are excited. E is excited in the course of receiving a signal from D. Intuitively, the excitation of B is not a cause of the excitation of E. However, the excitation of B strongly conditionally implies – in the sense of (SRT) – the excitation of E. For, once we have suspended judgement about the excitation of both B and E, the excitation of B lets us infer the excitation of C because there is no other way to excite B. Therefrom, we can infer that D fires, which in turn implies the excitation of E.

As observed by Paul and Hall (2013, p. 71-72), most counterfactual approaches solve the problem by excluding *backtracking* counterfactuals. We can adopt a similar strategy. The counterintuitive result about joint effects is avoided if all inferences that lead from the presumed cause C to the putative effect E are non-backtracking. We can make this idea precise in proof-theoretic terms. Recall that any inferential step in a natural deduction proof consists of a set P of premises and a conclusion C, where P may contain subproofs as premises. With this in mind, we can define the notion of a *forward-directed proof*. Such a proof conforms to the temporal order of events in the following sense:

Definition 5. $H \vdash_F C$
Let H be a set of formulas and C be a formula. Only literals and conjunctions of literals are taken to assert the occurrence of an event. We say there is a *forward-directed natural deduction proof* of C from H – in symbols $H \vdash_F C$ – iff there is a natural deduction proof of C from H such that for all inferential steps P/I (of the main proof and any subproof), if I asserts the occurrence of an event, then this event does not precede any event that is asserted by a premise in P or by a premise in a subproof that is a member of P.

The notion of an event is understood, in this definition, in the broad sense that includes negative events. A negative event is simply the failure of a corresponding positive event to occur. If A, B, or A and B are temporally extended events, we say that B does not precede A iff B does not come into being before A.

Furthermore, we require the derivation of the putative effect to be coherent in the following sense:

Definition 6. $H \vdash_K C$
Let H be a set of formulas and C be a formula. We say there is a *coherent natural deduction proof* of C from H – in symbols $H \vdash_K C$ – iff there is a natural deduction proof of C from H such that any assumption of any subproof is consistent with H. If a proof of C from H does not involve subproofs, this proof is vacuously coherent.

Why do we require that the assumption of any subproof is consistent with H? In our inferential approach to causation, subproofs are intended to represent possible ways how an event may actually bring about another event. Any assumption made in a subproof must therefore be consistent with the actual beliefs in H, the beliefs that form our premises. This condition will be relevant for the resolution of early and late preemption in sections 3.6 and 3.7. The condition is vacuously satisfied in all other causal scenarios considered here. We call the property in question *coherent* instead of *consistent* because it seems confusing to view certain correct natural deduction proofs as inconsistent.

For notational convenience, we introduce the notion of a forward-directed and coherent proof:

Definition 7. $H \vdash_{FK} C$
$H \vdash_{FK} C$ iff $H \vdash_F C$ and $H \vdash_K C$.

Using this notion of a forward-directed and coherent proof, we can impose further constraints on our Ramsey Test:

$$A \gg_{FK} C \in K(S) \text{ iff there is } \sigma \text{ s.t. } (K(S) \div_\sigma B(A) \vee B(C)), A \vdash_{FK} C. \quad (SRT_{FK})$$

That is, C is a forward-directed and coherent strong conditional implication of A iff we can suspend judgement on the antecedent A and the consequent C in such a manner that there is a forward-directed and coherent proof of C from A, in the context of the set of remaining beliefs of S. For C to be a cause of E, we require that C produces E, as expressed by the following condition:

$$C \gg_{FK} E \in K(S). \tag{C3}$$

Condition (C3) gives us the desired result about joint effects. To be more precise, we obtain the desired result if we represent the causal scenario of Figure 1 by the following prioritised belief base **H**:

$C \leftrightarrow D$, $D \leftrightarrow E$, $C \leftrightarrow B$
C, D, E, B

B cannot be a cause of E since there is no forward-directed and coherent natural deduction proof of E from B, in the context of the implications $C \leftrightarrow D$, $D \leftrightarrow E$, and $C \leftrightarrow B$. The requirement of using only forward-directed inferences is well motivated by our intuitions about causation: a cause must bring about, or produce, its effect. The inferential test of this production must be forward-directed in a manner that represents the temporal order of actual productive processes in the real world.

Note that a proof is forward-directed iff it is not backward-directed, i. e. does not involve inferences to the occurrence of an event that precedes any event asserted in the premises. Hence, a forward-directed proof may still involve inferences where the events asserted in the premises and the conclusion are simultaneous. The exclusion of backward-directed inferences will also prove crucial to solving the problem of preemption in sections 3.6 and 3.7.

A brief note on the bi-implications in the above belief base is in order. Why should we believe, e.g., $C \leftrightarrow B$ in place of merely believing $C \rightarrow B$? The above belief base assumes Figure 1 to represent a closed system. That is, there are no neurons other than *C* and *D* in play that could excite neurons *D*, *E*, and *B*. None of the general assertions about our analysis hinges on this assumption, though.

3.4 Overdetermination and Conjunctive Scenarios

The problem of overdetermination is severe for a counterfactual approach to causation in the tradition of Lewis (1973). This approach is centred on the idea of counterfactual dependence between cause and effect: if the cause had not occurred, the effect would not have occurred either – according to the simple counterfactual analysis. As is well known, this approach fails in cases described as *overdetermination*. Suppose a prisoner is shot by two soldiers at the same time, and each of the bullets is fatal without any temporal precedence. Then, intuitively, both shots would qualify as causes of the death of the prisoner. However, if one of the soldiers had not shot at the prisoner, the prisoner would still have died. The death of the prisoner does therefore not counterfactually depend on the shooting by a single soldier. Hence, on the counterfactual approach, neither of the soldiers is causally responsible for the death of the prisoner.

Or, consider the following neuron diagram:

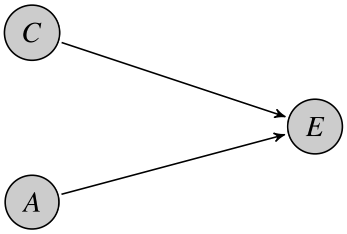

Figure 2

Neurons *C* and *A* fire so that neuron *E* is excited. The firing of each of *C* and *A* suffices to excite *E*. Hence, if *C* had not fired, *E* would still have been excited. This implies that, on the counterfactual approach, neither *C* nor *A* causes *E* to be excited. Such results are counterintuitive and not acceptable.

Our condition that $C \gg_{FK} E$ allows us to recognise relations of actual causation even in cases of overdetermination. It is easy to verify that the firing of *C* strongly conditionally implies the firing of E, in a forward-directed manner. In formal notation, $C \gg_{FK} E$, where C and E, respectively, stand for the firing of the neurons C and E. To verify $C \gg_{FK} E$, we need to suspend judgement about the firing of the neurons C and E. Suspending judgement about E forces us also to suspend judgement about A because our beliefs are inferentially closed. More precisely, if we were to retain the belief that A fires, we would also have to retain the belief that neuron E is excited because we believe that the firing of A triggers the firing of E. The latter belief has priority over beliefs about atomic events because it is intended to represent law-like connections between types of events. In sum, since A implies E and since we must retract our belief in E, we must also retract the belief in A.

In this way the suspension of judgement about C and E leads to an epistemic state in which we continue to believe the relations of firing between the neurons, but have no beliefs as regards the firing of C, A, and E. If we then assume that C is firing, we can infer that E is excited, in a forward-directed and coherent manner. Hence, $C \gg_{FK} E$. Thereby, we have epistemically verified that the firing of C produces the firing of E. These inferential considerations can easily be generalised so as to capture other examples, including those with fatal bullets.

We can therefore conclude that – thanks to (SRT_{FK}) – the problem of overdetermination does not arise in the first place for our analysis of causation. This is an important merit as compared to counterfactual approaches to causation. As we do not have to introduce further conditions to take overdeterministic causation into account, our analysis remains relatively simple and less likely to fall prey to further counterexamples that resemble scenarios of overdetermination.

Halpern and Pearl (2005) call the above neuron diagram a *conjunctive scenario* if the firing of both C and A is necessary to excite E. To give an example, it seems plausible that lightning together with a preceding drought is an – if not 'the' – actual cause of a forest fire. Indeed, it is again easy to verify that the firing of C strongly conditionally implies the firing of E, in a forward-directed and coherent manner. This time, the suspension of judgement regarding C and E does not force us to suspend judgement on A. The reason is that the firing of A alone is not sufficient to excite E. Hence, the suspension of judgement on C and E results in an epistemic state in which we continue to believe that A is firing. If we then assume that C is firing, we can infer that E is excited, in a forward-directed and coherent manner. The same reasoning applies to the verification of $A \gg_{FK} E$ due to the symmetry of the conjunctive scenario.

In a conjunctive scenario, where two events are necessary for an effect to occur, the conjunction of both necessary events should count as a cause. Our analysis captures this intuition by verifying $C \wedge A \gg_{FK} E$. While the definition of actual causation due to Halpern and Pearl (2005) recognises both events individually as causes, the conjunction is not recognised as such. They are forced to say that the lightning and the preceding drought together are no actual cause of the forest fire, while the lightning and the preceding drought individually are. Our analysis escapes such peculiarities that originate from Halpern and Pearl's formal apparatus.

3.5 Overdetermination Combined with a Conjunctive Scenario

In the causal scenarios considered so far, there has been only one way to suspend judgement as regards the presumed cause and the putative effect. In more technical terms, there has been but one epistemically optimal selection function with respect to the respective prioritised belief base. Hence, we could proceed without reference to different epistemically optimal selection functions. However, causal scenarios can be contrived such that there is more than one way to suspend judgement. Consider, for instance, the following neuron diagram:

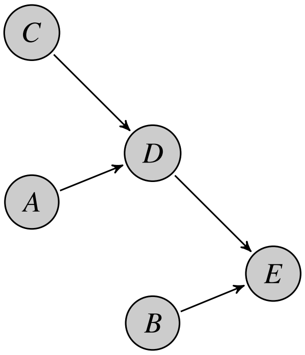

Figure 3

Suppose neurons C and A fire, and thereby overdetermine neuron D to be excited. But the excitation of both D and B is necessary to excite E. Figure 3 thus shows a combination of overdetermination with a conjunctive scenario. Here is a simple real-world example: suppose Billy is on vacation in Czech Republic. He has half a bottle of wine for dinner (C) and then two pints of beer in a pub (A). After the pub, he wants to drive home to his hotel. But Billy runs into a police control for alcohol (B). Clearly, his blood alcohol level is significantly above zero (D). Since Czech Republic has a zero tolerance policy about alcohol in the context of driving, Billy gets fined consequently (E). Obviously, the significant blood alcohol level is overdetermined by the wine and the beer. Moreover, the significant blood alcohol level and the alcohol test form two causes in a conjunctive scenario with the effect of Billy getting a fine. Finally, the event of drinking half a bottle of wine is clearly a cause of getting fined by the police. So, C is a cause of E. How does our approach fare with respect to this type of scenario?

There are two epistemically optimal ways to suspend judgment as regards C and E. First, A and D are retained, while B is given up alongside C and E. Second, B is retained, while A and D are given up alongside C and E. If we now assume C in the first case, we cannot infer E, since we cannot infer B and B is necessary for E. If we assume C in the second case, we can infer D. From D and the retained B, we can infer E. The constraints of forward directedness and coherence are satisfied in both cases. Hence, $C \gg_{FK} E$.

Let us verify the details by formalising the example. Here is the prioritised belief base $\mathbf{H}$ that represents the causal scenario:

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

Suspending judgement as regards C and E can be done in two ways. Why? Because there are two epistemically optimal selection functions with respect to $\mathbf{H}$. One yields H':

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

Why can we retain all these beliefs? Here is the diagram of a model that verifies all members of H' but fails to verify $C \vee E$:[3]

$$\{\neg C, A, D, \neg B, \neg E\}.$$

Hence, $C \vee E$ is not a logical consequence of H'. But any proper superset of H' that is a subset of $\bigcup \mathbf{H}$ entails $C \vee E$. Therefore, $H' \in \bigcup \mathbf{H} \perp C \vee E$. Since $H_1 \subseteq H'$, the

[3] A diagram of a model $\mathcal{A}$ is the set of all closed literals (in a given language) that are true in $\mathcal{A}$. The notion of a first order diagram has a clear analogue for propositional languages. In the case of propositional logic, a diagram contains for any propositional constant A, either A or $\neg A$. Such a diagram represents a valuation of a language of propositional logic.

selection function σ that picks H' is epistemically optimal. Hence, there is σ such that $\mathbf{H} \div_{\sigma} C \vee E = H'$. To carry out (SRT_{FK}), let us expand $K(H')$ by C. Obviously, $H', C \nvdash E$. Hence, σ does not confirm our causal intuitions in the present causal scenario. However, there is another epistemically optimal way of carrying out the agnostic move, yielding H'':

$$\boxed{(C \vee A) \leftrightarrow D, \quad (D \wedge B) \leftrightarrow E, \quad B}$$

Here is the diagram of a model that verifies all members of H'' but fails to verify $C \vee E$:

$$\{\neg C, \neg A, \neg D, B, \neg E\}.$$

By the same line of reasoning as for H', we can easily show that there is an epistemically optimal selection function function σ' such that $\mathbf{H} \div_{\sigma'} C \vee E = H''$. To carry out (SRT_{FK}), let us expand $K(H'')$ by C. Notably, $H'', C \vdash E$. There is even a forward-directed and coherent natural deduction proof of E from $\{C\} \cup H''$. σ' thus confirms our causal intuitions in the present scenario, while σ does not. By the definition of $\gg_{FK}$, this implies $C \gg_{FK} E$. Therefore, C does count as a cause of E.

The present example suggests that, for C to be a cause of E, it suffices that there is one agnostic context such that we can infer the putative effect from the presumed cause in a forward-directed and coherent manner. Figuratively speaking, the existential quantifier for σ lets us search for such a context. In a follow-up paper, we shall encounter further examples that require existential quantification over selection functions, for instance scenarios of double prevention.[4]

3.6 Early Preemption

Preemption is about backup processes: there is an event C that, intuitively, causes E. But even if C had not occurred, there is a backup event A that would have brought about E. Paul and Hall (2013, p. 75) take the following neuron diagram as canonical example of early preemption:

[4] Many thanks to an anonymous referee for pointing our attention to scenarios of the present type.

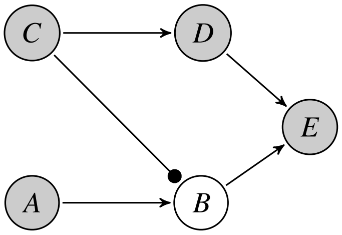

Figure 4

C's firing excites neuron D, which in turn leads to an excitation of neuron E. At the same time, C's firing inhibits the excitation of B. Had C not fired, however, A would have excited B, which in turn would have led to an excitation of E. Such scenarios are described as *early preemption* because the backup process does not get started in the first place, or is cut off prior to the occurrence of the effect.

How does our approach fare with respect to such scenarios? The good news is that we can easily, and without further modification, verify that C brings about E. That is, $C \gg_{FK} E$, where C and E, respectively, stand for the excitation of the neurons C and E. Once we have suspended judgement about C and E, the assumption C lets us infer, in the context of our background theory, that D and E. These inferences are forward-directed and coherent. The simple counterfactual approach, by contrast, fails here because it does not hold that the putative effect does not occur if the presumed cause is absent. In fact, E fires, even if C does not, due to the firing of A. This problem forced counterfactual approaches to be refined in various ways, see for example Lewis (1986), and Halpern and Pearl (2005).

$A \gg E$ holds as well, but the inferential path from A to E is not strictly forward-directed. Hence, $A \gg_{FK} E$ does not hold, which is more good news. To see this, let us suspend judgement as regards A and E. Clearly, the retraction of E requires the retraction of D and C. Otherwise, we could infer E using our beliefs about the relations of firing. These beliefs have priority over our beliefs about the firing of a neuron insofar as the relations of firing represent law-like connections between events. Trivially, retraction of A requires us to retract A. Notably, however, we do not retract $\neg B$ (the belief that B is not firing), as a consequence of retracting $A \vee E$. For, $\neg B$ neither implies A nor C, nor E. B may not fire because A does not fire or because C does fire, or both. After the agnostic move, we no longer know the reason why B does not fire. $\neg B$ merely implies $\neg A \vee C$, which does not suffice to establish C inferentially, as we shall see more clearly below.

The suspension of judgement on A and E results in an epistemic state in which we continue to believe $\neg B$, but have no beliefs about A, C, D, and E. If we now assume A, we can infer E, but only via the inferential path that goes through C and D. We cannot infer E in a forward-directed manner because we believe that B does not fire, and so we believe that B cannot excite E. The inferential path via C is backward-directed because it involves an inference from $\neg B$, A, and $(A \wedge \neg C) \leftrightarrow B$ to C. Clearly, this inference violates the condition of forward directedness as explained in Definition 5. For, from A firing and B not firing at a certain time t_B we can infer that C is firing at a certain time t_C, with the constraint that $t_C < t_B$. Hence, there is no forward-directed proof that would allow us to derive C from $\{\neg B, A, (A \wedge \neg C) \leftrightarrow B\}$. For simplicity, we leave the temporal order implicit, but it is obvious from the neuron diagram that a full explication of the temporal order would enable us formally to verify that the inference from A to E is not forward-directed.

It is worth formalising the example. This is the epistemic state **H** that represents the scenario:

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

Note that we do not only believe that $C \rightarrow D$ but also $D \rightarrow C$ because – in the confines of our causal scenario – D can only be excited by C. E, by contrast, can be excited by two different neurons. Hence, we do not believe that $E \rightarrow D$. We believe rather that E will not occur unless one of D or B does. Suspension of judgement as regards A and E results in an epistemic state H':

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

As explained above, it is obvious that we have to retract A, C, E, and D, as a consequence of retracting $A \vee E$. Why can we retain $\neg B$? Here is the diagram of a model that verifies all members of H' but fails to verify $A \vee E$:

$$\{\neg A, \neg B, \neg C, \neg D, \neg E\}.$$

Hence, $A \vee E$ is not a logical consequence of H'. Retaining $\neg B$ is even mandatory since H' is the only member of $\bigcup \mathbf{H} \bot (A \vee E)$ that respects the epistemic ranking of **H**.

To carry out (SRT_{FK}), let us expand $K(H')$ by A. Using $(A \wedge \neg C) \leftrightarrow B$ and $\neg B$, we can infer C from A, but not in a forward-directed manner, for the above explained reasons. In brief, C starts to fire before B is actually inhibited; the inhibition of B by C is not immediate. However, there is a forward-directed natural deduction proof of E from $\mathbf{H}'$ that makes use of reasoning by cases. Obviously, (i) we can prove $C \vee \neg C$ since this is a logical truth, even without backward-directed reasoning. Further, (ii) there is a forward-directed subproof of E from C. Likewise, (iii) there is a forward-directed subproof of E from $\neg C$ using A. By Disjunction Elimination, (i), (ii), and (iii) imply that there is a forward-directed proof of E from H'. Note, however, that the subproof of (iii) violates the coherence requirement of $\gg_{FK}$. For, in this subproof, we infer B from $\neg C$ and A, while $\neg B \in H'$. Since there is no other way to infer E from H', we conclude that $A \not\gg_{FK} E$. Hence, A does not count as a cause of E.

Note, finally, that we can generalise the argument that shows us why we retain certain beliefs about an intermediate event between the preempted cause A and the effect E, after the agnostic move as regards A and E has been carried out. Retaining such beliefs is feasible and mandatory on the following grounds: (i) The intermediate event in question does not occur because the causal process from A to E is cut off by the genuine cause C. (ii) Suspension of judgement concerning A and E requires us to suspend judgement about A, C, D and E, where D is an intermediate event between the genuine cause and the effect. (iii) The failure of the intermediate event B to occur may be due to A's failure to occur, the occurrence of C, or both. (iv) $\neg B$ does not allow us to infer the effect E via the preempted pathway. (ii) and (iii) imply that (v) we can neither infer C (which would imply D and E) nor $\neg A$ from $\neg B$. (v) and (iv) imply that there is no way to infer E from $\neg B$, once A, C, D, and E have been retracted. Hence, $\neg B$ is consistent with being agnostic about A and E. Since belief changes are to be as conservative as possible, it is mandatory to retain $\neg B$.

Most decisive and by no means trivial in this line of reasoning is (iii): the failure of the intermediate event to occur may be due to the non-occurrence of the preempted cause or the occurrence of the genuine cause, or both. As we are agnostic as to whether any of these causes occurs, we cannot infer from the non-occurrence of the intermediate event any claim as regards the occurrence of the genuine and the preempted cause. (If, by contrast, we were to believe that the preempted cause is present, we could infer the presence of the genuine cause from the non-occurrence of the intermediate event.) Hence, we can remain agnostic as regards the presence of the preempted and the genuine cause, while believing that the intermediate event does not occur.

3.7 Late Preemption

The canonical example of late preemption is presumably the most frequently cited causal scenario in the recent literature on actual causation. Suppose Billy and Suzy are throwing rocks at a bottle. Suzy's rock hits the bottle first, and thus is the genuine cause of the bottle's shattering. Billy, however, is also very skilful at throwing rocks. If Suzy had not thrown her rock, Billy's rock would have hit the bottle, and thus the bottle would have shattered a little bit later. Billy's throw is a preempted cause of the shattering of the bottle. Paul and Hall (2013, p. 99) describe this causal scenario as *late preemption*. In such a scenario, there is a backup process present that fails to go to completion merely because another process is effective prior to the backup one. In contrast to early preemption, the backup process is not cut off prior to the occurrence of the effect.

As is easy to verify, Suzy's throw strongly conditionally implies – in the sense of (SRT) – that the bottle shatters. Likewise, for Billy's throw. But only the first conditional implication can be established by a forward-directed and coherent proof. Hence, Billy's throw does not qualify as a genuine cause of the shattering of the bottle, as it should be.

Why is there no forward-directed proof of the shattering of the bottle from the assumption that Billy throws a rock at the bottle, once judgement has been suspended as regards the occurrence of these two events? The reasons for this are perfectly analogous to the reasons why there is no such proof in the case of early preempted causes. After judgement has been suspended, we continue to believe that Billy's rock did *not* hit the bottle. Therefore, there is no forward-directed inferential path from the assumption that Billy throws his rock to the shattering of the bottle. We can infer from Billy's throw and his rock not hitting the bottle that Suzy's rock has hit the bottle first. This inference, however, is not forward-directed because, by assumption, Suzy's rock arrives at the bottle before Billy's rock had a chance to hit.

Again, it is helpful to have a formal representation of the example:

ST: Suzy throws a rock at the bottle.

BT: Billy throws a rock at the bottle.

SH: Suzy's rock hits the bottle.

BH: Billy's rock hits the bottle.

BS: The bottle shatters.

We have adopted the symbols from Halpern and Pearl (2005), with the qualification that they stand for sentences instead of variables in a causal model. Here is the epistemic state $\mathbf{H}$ that represents the causal scenario:

$ST \leftrightarrow SH, \quad (SH \vee BH) \leftrightarrow BS, \quad (BT \wedge \neg SH) \leftrightarrow BH$
$ST, \quad BT, \quad SH, \quad \neg BH, \quad BS$

After the agnostic move as regards BT and BS, we obtain H':

$ST \leftrightarrow SH, \;\; (SH \vee BH) \leftrightarrow BS, \;\; (BT \wedge \neg SH) \leftrightarrow BH, \quad \neg BH$

Why do we continue to believe $\neg BH$ after judgement has been suspended as regards BT and BS? In other words, why is $BT \vee BS$ not a logical implication of $\bigcup H'$? Here is the diagram of a countermodel to this inference:

$$\{\neg ST, \neg BT, \neg SH, \neg BH, \neg BS\}.$$

It is easy to show that the model of this diagram verifies all members of $\bigcup H'$, while it does not verify $BT \vee BS$. Hence $BT \vee BS$ is not a logical implication of $\bigcup H'$.

Using $(BT \wedge \neg SH) \leftrightarrow BH$, we can infer from BT and $\neg BH$ that SH. This inference however is not forward-directed because the event asserted by SH precedes, at least slightly, the event asserted by $\neg BH$. (Recall that our notion of a forward-directed inference applies also to premises that assert the occurrence of a negative event, understood as the absence of a certain positive event.) In a manner analogous to the inferences in the above scenario of early preemption, we can infer BS from $H' \cup \{BT\}$ using reasoning by cases. But one of the subproofs violates the coherence requirement since it contains a line asserting BH. Since there is no other way to derive BS from $H' \cup \{BT\}$, we conclude $BT \not\gg_{FK} BS$. Billy's throw of the rock, though perfectly accurate, does therefore not qualify as a genuine cause of the shattering of the bottle. As is obvious, our solution exploits the consideration of *intermediate events* in similar ways as the solution by Halpern and Pearl (2005) does in the framework of structural equations.

Note once more that our insistence on forward-directed inferences is well motivated: we want to analyse causation as production by an epistemic verification that the presumed cause brings about the putative effect. Consequently, this epistemic verification must proceed by forward-directed inferences. Otherwise, our inferences would not model, or reconstruct, a process of production.

3.8 Switches

Switching scenarios are problematic for counterfactual accounts of causation. Lewis (1973), for example, defines actual causation as the transitive closure of causal dependence. If the distinct events C and E occur, then E causally depends on C just in case if C had not occurred, E would not have occurred. As a consequence, counterfactual dependence of E on C is sufficient for actual causation of the occurring event E by the occurring event C. This sufficiency for causation is widely shared among the counterfactual accounts in the tradition of Lewis, e.g. by Hitchcock (2001), Woodward (2003), Hall (2004), Hall (2007), and Halpern and Pearl (2005).

The following neuron diagram represents a simplified version of a switching scenario:

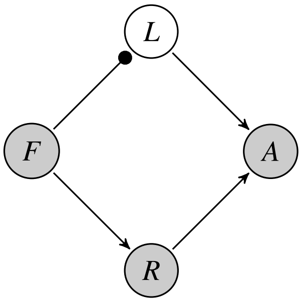

Figure 5

The firing of neuron F excites R's firing, which in turn excites neuron A. At the same time, F's firing inhibits the excitation of L, which would have been excited in case F had not fired. In brief, F determines which one of L and R is firing, and thus acts like a switch. Speaking of events, F, R and A occur, and both R counterfactually depends on F and A counterfactually depends on R. (The latter counterfactual dependence holds on the assumption that counterfactuals are non-backtracking.) By the transitive closure imposed on the one-step causal dependences, Lewis (1973) is forced to say that F is a cause of A.

The counterfactual accounts based on structural equations due to Hitchcock (2001) and Halpern and Pearl (2005) reject the transitivity of causation. Still, Hitchcock (2001) counts F to be a cause of A. The reason is that there is an active path from F over R to A and keeping the off-path variable L fixed at its actual value induces a counterfactual dependence of A on F. Similarly, Halpern and Pearl (2005) count F to be a cause of A, since A counterfactually depends on F under the contingency that $\neg L$.

Hall (2007, p. 28) puts forth a switching scenario in which F should intuitively not count as a cause of A: Flipper is standing by a switch in the railroad tracks. A train approaches in the distance. She flips the switch (F), so that the train travels down the right-hand track (R), instead of the left (L). Since the tracks reconverge up ahead, the train arrives at its destination all the same (A).

Flipping the switch does not seem to be a cause of the train's arrival. By assumption, 'the train arrives at its destination all the same' independent of the flipping. Hence, flipping the switch makes no difference to the train's arrival. And unlike scenarios of preemption, the lack of a net effect is not due to a backup process.

Our analysis verifies that F brings about A. Once we have suspended judgement on F and A, the assumption of F lets us infer – in a forward-directed and coherent manner – that R and thus A. Here, we see that production is necessary but not sufficient for causation. Although F is a producer of A, F is not a cause of A. Our diagnosis of the problem is: while it is possible that F brings about A and the absence of F would also bring about A, in the sense of (SRT_{FK}), it sounds paradoxical to say that F causes A and the absence of F would cause A as well. The concept of causation seems to preclude that both the presence and absence of an event can genuinely cause the same effect. Similarly, Sartorio (2005, p. 90) states "if causes are difference-makers, it is in virtue of the fact that events and their absences would not have caused the same effects."

Therefore, we complement our analysis by a weak condition of difference-making. For C to be a cause of E, the absence of C does not strongly conditionally imply

E, in a forward-directed and coherent manner:

$$\neg C \gg_{FK} E \notin K(S). \tag{C4}$$

Condition (C4) says that an event is only an actual cause if its absence does not also bring about the effect. Importantly, this condition is weaker than the difference-making in counterfactual approaches as it does not require the putative effect to be absent if the presumed cause fails to occur. The condition merely demands that the putative effect cannot be inferred from the absence of the presumed cause, once the agnostic move has been made.[5]

Let us now return to the switching scenario. We know already that F brings about A. That is, $F \gg_{FK} A$. However, the absence of F would also bring about A, in a forward-directed and coherent manner. Hence, $\neg F \gg_{FK} A$. Flipping the switch and not flipping the switch individually bring about the arrival of the train. Once Condition (C4) is in place, however, flipping the switch is no cause of the train's arrival. However, our analysis verifies that F causes R and R causes A. Here, the transitivity of causation fails because we can infer too much: the effect A will trivially obtain independent of whether or not F occurs.

To ease the verification of the details, we explicate the epistemic state $\mathbf{H}$ that represents the causal scenario:

$F \leftrightarrow R, \quad \neg F \leftrightarrow L, \quad (R \vee L) \leftrightarrow A$
$F, \quad R, \quad A, \quad \neg L$

After the agnostic move as regards F and A, we have the epistemic state H':

$F \leftrightarrow R, \quad \neg F \leftrightarrow L, \quad (R \vee L) \leftrightarrow A$

If we now assume F, we obtain R and thus A. The assumption of $\neg F$ results in L and thus A.

Unlike the standard counterfactual condition of difference-making, (C4) does not imply counterintuitive results about overdetermination. Let C be a putative cause of an effect E in a scenario of overdetermination. If we test for $\neg C \gg_{FK} E \notin K(S)$, the agnostic move forces us to suspend judgment about the occurrence of

[5] Condition (C4) is inspired by Rott (1986). Recently, a conceptually similar condition has been used by Beckers and Vennekens (2017, 2018) in an analysis of actual causation.

all overdetermining causes. (The reason is that we can infer the effect from any overdetermining cause.) Hence, if we assume – after the agnostic move – that one of the overdetermining causes is absent, we are unable to infer the effect. $\neg C \gg_{FK} E \notin K(S)$ is therefore satisfied for causal relations in scenarios of overdetermination. Similar considerations apply to conjunctive scenarios and cases of preemption.

To sum up, our forward-directed and coherent conditional $\gg_{FK}$ verifies some cases where the antecedent does not count as a cause of the consequent in any reasonable sense. Hence, we have complemented Condition (C3) by the weak difference-making Condition (C4) to yield a more appropriate account of causation in terms of production. As an upshot, our analysis of causation combines a notion of production and a weak notion of counterfactual dependence. Here our analysis stands in stark contrast to the postulate of Hall (2004) that there be two stand-alone concepts of causation. We rather agree with Hall (2007) that production and (weak) counterfactual dependence are metaphysically entangled in *one* concept of causation.

3.9 Joint Effects Revisited: Spurious Causes

In Section 3.3, we dealt with the problem of joint effects using the notion of a forward-directed (and coherent) proof. This solution works on the tacit assumption that our inferences track the presumed causal paths, which are illustrated by the arrows in the neuron diagrams. For this to be seen, let us revisit the neuron diagram of joint effects:

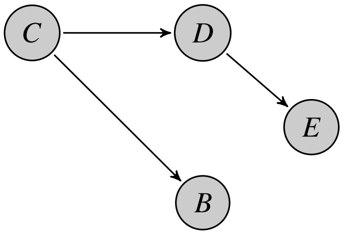

Figure 1

In the above analysis of joint effects, our inferences stay on causal paths since it was assumed that any implication and bi-implication in the belief base $H = \bigcup \mathbf{H}$

corresponds to an elementary causal path, while the bi-implication $B \leftrightarrow E$ is not a member of H. However, as there is a stable correlation between the firing of B and E, it is rather plausible to have $B \leftrightarrow E$ in the belief base. So let $H_1' = \{C \leftrightarrow D, D \leftrightarrow E, C \leftrightarrow B, B \leftrightarrow E\}$, where $\mathbf{H}' = \langle H_1', H_2 \rangle$. Unfortunately, B is a cause of E relative to the epistemic state $\mathbf{H}'$, on our present analysis. It is easy to verify that $B \gg E$ with respect to $\mathbf{H}'$. If the firing of B precedes that of E, there is a forward-directed inferential path from B to E, and so $B \gg_F E$ holds as well. The other conditions of our analysis are satisfied in an obvious way. Hence, $B \gg_{FK} E$. So we would have to consider the firing of B as a cause of the firing of E after all. This is counterintuitive. The underlying problem is, that once we adopt the bi-implication into the belief base, our forward-directed inferences may leave the presumed causal paths.

On the tacit assumption that our inferences stay on the presumed causal paths, the problem of joint effects merely requires to block backward-directed inferences from certain effects to their common causes. If, by contrast, our inferences leave the presumed causal paths, the problem of spurious causation arises. We take an event A to be a spurious cause of an event D if D can be inferred from A, in a forward-directed manner, but the inference does not represent a presumed causal path. In what follows, we give a brief summary of an account of spurious causation that has been developed in greater detail in Andreas (2019). The basic idea is that forward-directed inferences with a genuine causal meaning must be based on non-redundant generalizations.

So far, we have represented the dynamics of causal scenarios by very simple implications, without attempting to spell out a law-like generalisation from which these implications can be derived. We can represent the dynamics of neuron diagrams in a more fine-grained way by making the following concepts explicit:

- $AC(x, y)$: there is a direct neuronal connection between x and y, and activation of x is followed by activation of y.
- $A(x)$: neuron x is active.

These concepts at hand, we can recognize a general law governing the behaviour of neurons in simple neuron diagrams:

$$\forall x \forall y (AC(x, y) \land A(x) \to A(y)) \tag{γ_1}$$

While noting that γ_1 could easily be modified to take inhibitions into account, we leave them out for simplicity. Further, let us represent the neurons of Figure 1 by the constants c, d, e, and b, where these constants have their obvious meanings.

Recall that the problem of joint effects arises only if we interpret Figure 1 in the sense of a closed system. That is, there are no neurons in play other than the ones depicted in this figure. Also, we have tacitly assumed that Figure 1 gives us a complete representation of the connections between the neurons depicted. Otherwise, the belief base **H** in Section 3.3 would not correctly represent the dynamics of the causal scenario under consideration. Similar considerations apply to the problem of spurious causation. If Figure 1 does not represent a closed system, the problem of spurious causation does not arise in the first place. For, the sentences $A(b) \rightarrow A(e)$ and $A(b) \leftrightarrow A(e)$ may well be false then; we would not be justified in accepting these sentences. In Section 3.3, we have made the closed-system assumption explicit using simple bi-implications. If, by contrast, we want to derive assertions about the dynamic behaviour from a proper generalization, the sentential representation of the closed-system assumption is a bit more involved:

$$\forall x \forall y (AC(x,y) \land A(y) \land \neg \exists z (A(z) \land z \neq x \land AC(z,y)) \rightarrow A(x)) \tag{γ_2}$$

$$\forall x (AC(x,d) \rightarrow x = c) \tag{α_1}$$

$$\forall x (AC(x,e) \rightarrow x = d) \tag{α_2}$$

$$\forall x (AC(x,b) \rightarrow x = c) \tag{α_3}$$

α_1, α_2, and α_3 resist a straightforward classification into our two epistemic levels $G \cup I$ and F. For one thing, they have the logical form of universal propositions, and for another, they are shorthand for asserting certain negative facts. The good news, however, is that the following analysis of spurious causation does not hinge at all on the classification of α_1, α_2, and α_3 into whatever epistemic levels. To avoid confusion, we put these sentences into a separate epistemic level. Now we are in a position to represent the causal scenario in a more fine-grained way by a prioritised belief base $\mathbf{H}'$:

γ_1, γ_2
$\alpha_1, \alpha_2, \alpha_3$
$AC(c,d)$, $AC(d,e)$, $AC(c,b)$, $A(c)$, $A(d)$, $A(e)$, $A(b)$

Note that the sentences $AC(c,d), AC(d,e)$, etc. are atomic sentences. Hence, they represent statements about atomic facts. At the same time, these sentences encode, in the context of γ_1, information about the dynamic behaviour of neurons. It is open to interpretation whether or not sentences of the form $AC(c_1, c_2)$ encode primitive causal relations. Note that the meaning of such sentences also depends on the precise understanding of neuron diagrams. There are at least two interpretations of the arrows in such a diagram: (i) as encoding primitive causal relations and (ii) as actual connections of a physical system within an oversimplified account of neuronal systems. We adopt the latter interpretation in this section since we merely want to give an example that motivates a general account of spurious causation. The general account may or may not enable us to give a reductive analysis of causation in the long run. For the time being, we prefer to be as neutral as possible on this question.[6]

How does the more fine-grained representation help resolve the problem of spurious causes? What is the decisive difference between $\mathbf{H}'$ and $\mathbf{H}$? If we extend $\mathbf{H}'$ by $A(b) \leftrightarrow A(e)$, $A(b) \gg_{FK} A(e)$ becomes a member of $\mathbf{H}'$ thus extended. Both $A(b) \leftrightarrow A(e)$ and $B \leftrightarrow E$ are redundant in their respective belief base. So, the logical behaviour of $\mathbf{H}'$ and $\mathbf{H}$ seems almost completely analogous as regards the causal connection in question.

Nonetheless, there is an important difference between $\mathbf{H}'$ and $\mathbf{H}$. In the case of $\mathbf{H}$, all generalisations of H_1 are redundant in the set $H_1 \cup H_2 \cup \{B \leftrightarrow E\}$. In the case of $\mathbf{H}'$, by contrast, $A(b) \leftrightarrow A(e)$ is the only member of H_1' that is redundant in the set $H_1' \cup H_2' \cup H_3' \cup \{A(b) \leftrightarrow A(e)\}$, where $H_1' \cup H_2' \cup H_3' = \bigcup \mathbf{H}'$. Hence, we can break the symmetry between genuine and spurious causal relations by imposing another constraint on the inferential path between the presumed cause and the putative effect:

Definition 8. $H \vdash_N C$
Let H be a set of formulas and C be a formula. We say there is a natural deduction proof of C from H based on non-redundant generalisations and implications – in symbols $H \vdash_N C$ – iff there is a natural deduction proof of C from H such that any formula ϕ subsequent to the premises has the following property: if ϕ is an implication or generalisation, ϕ is non-redundant in H.

Definition 9. Redundancy
A sentence $\phi \in A$ is redundant in A – relative to a logic $\models$ – iff $A \setminus \{\phi\} \models \phi$.

[6] In Andreas and Günther (2018, eprint), we defined the strengthened Ramsey Test in the framework of causal models by Halpern and Pearl (2005). Thereby, we put forth a variant of the present analysis of actual causation in terms of causal models. This analysis is not reductive since structural equations of a causal model are presumed to encode primitive causal relations.

The following strengthening of $\gg_{FK}$ now suggests itself:

$$\begin{aligned} A \gg_{FKN} C \in K(S) \ \text{ iff there is } \sigma \text{ s.t.} \\ (K(S) \div_{\sigma} B(A) \lor B(C)), A \vdash_{FKN} C. \end{aligned} \qquad (SRT_{FKN})$$

$\vdash_{FKN}$ has its obvious meaning. Now, we require there to be a forward-directed and coherent proof of the putative effect from the presumed cause such that any generalisation and implication of this proof is non-redundant. That is, we rewrite condition (C3) as follows:

$$C \gg_{FKN} E \in K(S). \qquad \text{(C3)}$$

This condition solves the problem of spurious causation, with the qualification that very simple representations of causal scenarios may not suffice to properly discriminate between genuine and spurious causes. Two different representations of one and the same causal scenario may well yield different causal verdicts. If there is such a difference, one should go for the more fine-grained representation. Our final analysis will be explicitly relativised to an epistemic state S. This is comparable to Halpern and Pearl's (2005) analysis being relative to a causal model. Some epistemic states are just too simplistic for a proper recognition of causal relations.

As is the case with **H** extended by the bi-implication $B \leftrightarrow E$, there are even sets of implications where all implications are redundant so that no causes can be recognised using the conditional $\gg_{FKN}$. However, once we work out a more fine-grained representation using full first-order logic and, ideally, some scientific knowledge, the problem can be addressed in a satisfying manner.

Is there any heuristics for devising "good" propositional representations of causal scenarios? For this, we suggest imposing the following constraint on the implications in a belief base: any implication $A \rightarrow B$ must be derivable from non-redundant generalizations without this derivation involving inferential paths via events that are part of the propositional representation, but different from A and B. The implication $B \rightarrow E$, for example, does not meet this condition in the above causal scenario of joint effects. Likewise, the implication $BT \rightarrow BS$ fails to meet the condition in the Suzy-Billy scenario. Note, however, that the present heuristics rests on more fine-grained representations of causal scenarios using first-order logic and, ideally, some scientific knowledge.

What is the philosophical motivation for the requirement that the inferential path from a genuine cause to its effect must be based on non-redundant generalisations?

As shown in Andreas (2019), our analysis of causation may easily be extended to an inferential approach to causal explanation. Drawing on a *best system account* of laws of nature, we claim that a proper inferential explanation is based on non-redundant generalisations. Further, we conjecture that in the case of deterministic spurious causation our intuitions about causal relations are constrained by our intuitions about proper inferential explanations. This conjecture has been validated for real-world examples in Andreas (2019).

4 Conclusion

It is time for a concluding summary of our analysis. We represent causal scenarios by epistemic states, which in turn are given by prioritised belief bases. We define:

Definition 10. ***C*** **is a cause of** ***E***
The event (asserted by) C is a cause of the event (asserted by) E – relative to an epistemic state S – iff

(C1) $C, E \in K(S)$,

(C2) $t(C) < t(E)$,

(C3) $C \gg_{FKN} E \in K(S)$, and

(C4) $\neg C \gg_{FKN} E \notin K(S)$.

Definition 10 puts forth an analysis of causation in terms of a strengthened Ramsey Test conditional, which is meant to express a relation of production. By condition (C3), production is necessary but not sufficient for actual causation. Similarly, (C4) amounts to a weak condition of difference-making that is necessary but not sufficient for actual causation. A cause is a difference-maker in the weak sense that its presence and its absence cannot bring about the same effect. From this follows the principle of difference-making convincingly argued for by Sartorio (2005): if C is a cause of E, $\neg C$ is not a cause of E. Note that we can derive this principle from (C3) and (C4). Suppose C is a cause of E. Hence, by (C4), $\neg C \gg_{FKN} E \notin K(S)$. By (C3), this implies that $\neg C$ is not a cause of E. Notably, Sartorio views the principle in question as a condition “that the true analysis of causation (if there is such a thing) would have to meet” (ibid., p. 75).

The present analysis is an epistemic approach to causation in the tradition of Hume, Gärdenfors (1988, Ch. 9), and Spohn (2006). The following objection to this line of research suggests itself: a Ramsey Test analysis of causation merely explicates an epistemic notion of causation, while it leaves the metaphysical notion of causation unexplained. We shall address the apparent tension between the metaphysical and the epistemic notion of causation elsewhere. Likewise, we shall deal with scenarios of prevention and double prevention in a follow-up paper. However, in this paper we have achieved a formally elaborated solution to the problems of overdetermination, conjunctive scenarios, (early and late) preemption, and switches. To the best of our knowledge, there is no other formally well-defined account in the literature that solves the whole set of these problems.

Acknowledgments. We would like to thank Hannes Leitgeb and Andrew Irvine for very valuable advice on earlier versions of the paper. Special thanks are also due to Hans Rott and Paul Bartha for very helpful comments on presentations of the paper. Finally, we are indebted to the anonymous referees for Philosophical Studies. Their comments greatly helped us improve the paper. This research has been supported, in part, by the Graduate School of Systemic Neurosciences.

References

Alchourrón, M. A., Gärdenfors, P., and Makinson, D. On the Logic of Theory Change: Partial Meet Contraction Functions and Their Associated Revision Functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

Andreas, H. Explanatory Conditionals. *Philosophy of Science*, 86(5), 2019.

Andreas, H. and Günther, M. On the Ramsey Test Analysis of 'Because'. *Erkenntnis*, 2018.

Andreas, H. and Günther, M. A Ramsey Test Analysis of Causation for Causal Models. *The British Journal for the Philosophy of Science*, 2018, eprint.

Beckers, S. and Vennekens, J. The Transitivity and Asymmetry of Actual Causation. *Ergo: An Open Access Journal of Philosophy*, 4:1–27, 2017.

Beckers, S. and Vennekens, J. A Principled Approach to Defining Actual Causation. *Synthese*, 195(2):835–862, Feb 2018.

Bradley, R. A Defence of the Ramsey Test. *Mind*, 116(461):1–21, 2007.

Gärdenfors, P. *Knowledge in Flux.* MIT Press, Cambridge, MA, 1988.

Gärdenfors, P. Conditionals and Changes of Belief. In Niiniluoto, I. and Tuomela, R., editors, *The Logic and Epistemology of Scientific Change*, volume 30 of *Acta Philosophica Fennica*, pages 381–404. 1978.

Gärdenfors, P. Belief Revisions and the Ramsey Test for Conditionals. *The Philosophical Review*, 95(1):81–93, 1986.

Hall, N. Structural Equations and Causation. *Philosophical Studies*, 132(1):109–136, 2007.

Hall, N. Two Concepts of Causation. In John Collins, Ned Hall, and Laurie Paul, editors, *Causation and Counterfactuals*, pages 225–276. The MIT Press, 2004.

Halpern, J. Y. and Pearl, J. Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science*, 56(4): 843–887, 2005.

Hansson, S. O. *A Textbook of Belief Dynamics. Theory Change and Database Updating*. Kluwer, Dordrecht, 1999.

Hansson, S. O. In Defense of the Ramsey Test. *Journal of Philosophy*, 89(10): 522–540, 1992.

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

Hume, D. *A Treatise of Human Nature*. Oxford: Clarendon Press, 1739/1978.

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

Lewis, D. Postscripts to "Causation". In Lewis, D., editor, *Philosophical Papers. Volume II*, pages 172–213. Oxford University Press, Oxford, 1986.

Paul, L. A. and Hall, N. *Causation: A User's Guide*. Oxford University Press, 2013.

Ramsey, F. P. General Propositions and Causality. *Philosophical Papers*, pages 145–163, 1929/1990.

Rott, H. Ifs, Though, and Because. *Erkenntnis*, 25(3):345–370, 1986.

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

Spohn, W. Causation: An Alternative. *British Journal for the Philosophy of Science*, 57(1):93–119, 2006.

Stalnaker, R. A Theory of Conditionals. In Rescher, N., editor, *Studies in Logical Theory (American Philosophical Quarterly Monograph Series)*, number 2, pages 98–112. Blackwell, Oxford, 1968.

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