THE REVIEW OF SYMBOLIC LOGIC
Volume 6, Number 1, March 2013

# ON FORMALIZING CAUSATION BASED ON CONSTANT CONJUNCTION THEORY

HU LIU and XUEFENG WEN
Institute of Logic and Cognition, Sun Yat-sen University, China

**Abstract.** Constant conjunction theory of causation had been the dominant theory in philosophy for a long time and regained attention recently. This paper gives a logical framework of causation based on the theory. The basic idea is that causal statements are empirical, and are derived from our past experience by observing constant conjunction between objects. The logic is defined on linear time structures. A causal statement is evaluated at time points, such that its value depends on what has been in the past. We first give a semantics that contains basic conditions that, we think, must hold for a concept of causation, on which we define the minimal causal logic. Then we discuss its possible extensions for various concepts of causation. Complete deductive systems are given.

**§1. Introduction.** The ability of building causal relations is fundamental for us in understanding the world in everyday life. Not surprisingly, causation has been one of the core concepts in philosophical study since antiquity. Since formal logic was founded, logicians have put a number of philosophical concepts under formal investigations, yielding various branches of nonclassical logic. Causation has also been formalized in logic (Lewis, 1973a; von Kutschera, 1993; Lin, 1996; Xu, 1997; McCain & Turner, 1997; Thielscher, 1997; Schwind, 1999; Giordano *et al*., 2000; Pearl, 2000; Zhang & Foo, 2001; Giordano & Schwind, 2004; Halpern & Pearl, 2005). As far as we know, however, no work has been done in formalizing the concept of causation based on constant conjunction theory (a.k.a. regularity theory). In this paper, we formalize constant conjunction and causation with the help of a linear temporal logic.

Constant conjunction theory (CCT, henceforth) had been the dominant theory in philosophy for a long time. It explains causation with a central pretheoretic intuition that an effect is something that always accompanies its cause. Unsurprisingly, it still has found support in scientific discoveries, such as Pavlov's laws of conditioning and Hebbian neural networks (Hebb, 2002). However, CCT has virtually vanished from philosophical discussions in decades, until very recently, some philosophers have directed their attention back to this account of causation. See, for example, Graßhoff & May (2001), Hall (2004), and Beebee (2006), and also Mackie (1974) for an earlier attempt. See also Baumgartner (2008) for replies to criticisms against CCT, and Baumgartner (2009) for an application of CCT in a coincidence analysis in deterministic causal structures. As argued by Baumgartner (2008), there are several advantages of CCT over other theories of causation:

> [U]nlike e.g. counterfactual accounts regularity theories straightforwardly handle cases of overdetermination. As against interventionist or manipulatory accounts, analyses of causation in terms of regularities do not run

Received: September 18, 2012.

© Association for Symbolic Logic, 2012

doi:10.1017/S1755020312000299

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

> the risk of being anthropocentric. Contrary to probabilistic accounts, regularity theories are not compromised by paradoxical data as, for instance, generated in cases of Simpsons Paradox. Finally, while transference theories treat a fundamental type of causal process, transference processes, as conceptually primitive and thus do not attempt to provide a reductive analysis of causation, regularity accounts, properly conceived, offer the promising prospect of explicating the cause-effect relation C or at least one among possibly several types of cause-effect relations C in entirely non-causal terms.

Unlike its overwhelming opponent, the so-called counterfactual theory whose logical frameworks have been well-studied (Lewis, 1973a), logical theories of CCT have been missed in literature. It is our purpose of this paper to start such researches.

Why is the formalization of CCT interesting? Firstly, unlike other theories of causation, like counterfactual theory, CCT is an extensional theory. Hence, it is more natural to be embedded in the standard extensional logical framework (augmented with a time structure). A logic of causation based on CCT need not assume metaphysical concepts like possible worlds and similarity between possible worlds, which are usually required in the logic of causation based on counterfactual theory. These metaphysical concepts are highly controversial.

Secondly, with a logic for CCT in hand, we can get more explicit properties and consequences of CCT and thus can evaluate the theory better. Moreover, with a less metaphysically burdened framework, we can formalize many other theories of causation in it. In this way, we are hopeful to get deeper understanding of these theories and thus the notion of causation.

There are plenty of controversies about causation in philosophy. The present paper concerns a formal framework for an existing philosophical theory. We will try to be involved as little as possible in philosophical debate. For this purpose, we shall first settle several conceptual bases that are assumed in the paper.

We follow the standard view that causation is a relation between two objects, one called cause and the other effect. We introduce a binary causal operator $C$ to formalize this relation. A formula $C(\alpha, \beta)$ means that $\alpha$ causes $\beta$.

Another standard view of causation is that objects related as causes and effects are events (Lewis, 1986). There is also considerable support for facts instead of events as the causal relata (Mellor, 1995). We will take the event view in our presentation, where propositional variables represent atomic events, and arbitrary formulas represent complex events. For this, we assume that there exist certain atomic events, based on which complex events are built by connectives $\neg$, $\vee$, and so on. It would make no difference if we understand the logic as a causal logic for facts. The formal framework is the same.

Many philosophers, as well as most ordinary persons, would agree that causation is closely related to the concept of time, in that a cause is no later than its effects. This is taken to be a distinction between general conditionals and causal conditionals, for a general conditional statement does not require a temporal order. Some philosophers, including Kant (Kant, 1933)-A203, argued for simultaneous causation. See Mellor (1995) for more details and replies to simultaneous causation. There has been a relatively new debate in philosophy about backward causation, supported by Dummett (Dummett & Flew, 1954) and rebutted by Flew (Dummett & Flew, 1954) and Black (Black, 1956).

We follow the line of epistemic causation, or psychological causation (Williamson, 2006), which started from Hume, and had been the dominant theory until decades ago.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

Hume stated his constant conjunction explanation of causation with the famous first definition: "We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second." (Hume, 2000, sec. VII). According to Hume, causal statements are empirical, and the idea of causation is derived from human's experience. Causation is just a simplification for the convenience of the mind. CCT opposes both simultaneous causation and backward causation. Thus, we assume a cause always precedes its effects.

Hume's definition is for causation of particular events. A causal relation between $\alpha$ and $\beta$ is derived from the fact that $\beta$ did follow $\alpha$, and all occurrences of events of type $\alpha$ (or in Hume's words, events similar to $\alpha$) were followed by occurrences of events of type $\beta$ (events similar to $\beta$). The present logic is meant to apply to causal generalizations, not to causation in particular cases. We deal with general causal statements like "events of type $\alpha$ cause events of type $\beta$."

For causal generalizations, the second part of Hume's definition suffices. We say that $\alpha$ *constantly conjoins* with $\beta$, if we have observed that after each occurrence of an $\alpha$ event, an occurrence of a $\beta$ event was present. Each such pair of occurrences of events is an evidence that $\alpha$ constantly conjoins with $\beta$ (also an evidence that $\alpha$ causes $\beta$). We will simply say that $\alpha$ occurs at $t$, or $\alpha$ is true at $t$, with the intended meaning that an event of type $\alpha$ occurs at $t$.

Let $t < u$ mean that time point $t$ precedes time point $u$. Suppose that $\alpha$ occurred at $t$, and $\beta$ occurred at $u$ with $t < u$. Clearly, for this pair of occurrences of $\alpha$ and $\beta$ to be a legitimate evidence of $\alpha$ causing $\beta$, there should be a constraint on the length of time between $t$ and $u$. Even though we have observed that after each thunder, a year later it snows, we would not build a causal relation that thundering causes snowing. The time period is too long for a thunder to be influential. The principle of closeness, or contiguity, between cause and effect is an indispensable part of CCT, and is also crucial in a formal framework for this type of causation.

The question is how long it should be for a pair of occurrences of events to be contiguous, and therefore a legitimate evidence. This measurement has not been a subject in philosophy. It should not depend on effects. It is a measurement that how long the influence of a cause lasts. A massive event such as an earthquake should have a longer *effect period* than a minor event such as I eating my breakfast. Different causes may have different lengths of influence. This is almost all we can say about the measurement. There shall be no systematical way to determine an effect period for each object, partially because it is a matter of our psychological perspectives of objects.

We will not concern ourselves with the lengths of effect periods in this paper. We will work with a general linear time structure, in which for each event $\alpha$, its effect period is modeled by a binary relation $R_\alpha$. At time point $t$, those points in $R_\alpha(t) = \{u \mid R_\alpha tu\}$ are considered contiguous to $t$ with respect to $\alpha$. Any $\beta$ that occurs at some point in $R_\alpha(t)$ is considered contiguous to the occurrence of $\alpha$ at $t$. An arbitrary binary relation is not suitable for the concept of effect period. We shall impose the following restrictions on $R_\alpha$:

(1) A cause can only affect the future, which requires that $R_\alpha tu$ implies $t < u$.
(2) A cause's influence lasts for a period, that is, time points under the influence of an occurrence of $\alpha$ should be uninterrupted, which requires that if $R_\alpha tu$ and $t < s < u$ then $R_\alpha ts$.

Causation is not guaranteed by constant conjunction. There are plenty of examples that two objects are constantly conjoined, whereas they are not causally linked. For the

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

simplest example, consider a tautology $\alpha$ (or anything having always been true in the past), which by no mean can be a cause or an effect. However, if some $\beta$ with a nonempty effect period happened in the past, $\alpha$ would certainly occur in the effect periods of every instance of occurrence of $\beta$. We cannot take this constant conjunction as a causation. Something should be done in distinguishing causal relations from those noncausal constant conjunction relations. This, however, is where a philosophical debate is inevitable. We will try to avoid the disputation by only setting the minimal conditions that a concept of causation should satisfy (see the next section). The corresponding logic will be called the *minimal causal logic*. We will discuss other plausible conditions in Section 3, which correspond to extensions of the minimal logic.

Our logic is a direct formalization of the constant conjunction explanation of causation. All causal relations are directly built from the observations of the past. The real process, however, may be much more complicated. For example, as the famous idea of theory-laden observation suggested (Hanson, 1958), theories are indispensable part of observations. A real process may involve the influence of the background knowledge. This, or other types of complications, are not discussed in this paper.

Our logic is built on linear time structures. Basically, there are two types of logics of time, first-order (or high-order) style and modal style. Both are reasonable ways to handle time and they are closely related to each other. The first-order approach is in general more expressive and more succinct. The modal approach has caught much attention. It is better in computational cost, and arguably closer to natural language. The first-order approach is external: sentences of the logic are evaluated with respect to whole structures. The modal approach is internal: formulas are evaluated at individual time points. We prefer the modal approach not only because of its popularity, but also because an internal perspective is more suitable for our purpose. The truth value of a causal statement $C(\alpha, \beta)$ depends on what have happened in the past of the *current* point. The evaluation has to be at a given current point. See Hodkinson & Reynolds (2007) for more discussion on the two approaches.

Constant conjunction is a type of learning mechanism. Machine learning has been intensively studied in AI with quite different views and techniques. These are algorithmic rather than logical. However, they may share some motivation with the current causal logic. The closest area to the current one is the so-called Reinforcement Learning (Sutton, 2001), where algorithms are designed to learn the effects of actions from observations, so as to take "right" actions to maximize some notion of rewards. Its focus is on finding a balance between exploration (observations) and exploitation (current knowledge). In this paper, we assume that observation is always prior to the current knowledge, which means that we unconditionally change our causal knowledge according to what we have observed. A trade-off between knowledge and observation would be an interesting topic in future research. Techniques in reinforcement learning may help for this purpose.

The remainder of the paper is structured as follows. The next section presents the minimal causal logic, including its axiomatization. Section 3 is on various extensions of the minimal causal logic by imposing conditions on either models or semantics. Also, we discuss the transitivity of causation and the inus condition in this section. In Section 4 we compare the current logic with existing causal logics. We conclude the paper by suggesting some future work.

## §2. The minimal causal logic.

***2.1. Syntax.*** The alphabet of *causal logic* (CL) consists of a countable set $P = \{p, q, \ldots\}$ of propositional variables, Boolean connectives $\neg$ and $\vee$, temporal

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

operators $G$ and $H$, a binary causal operator $C$, and effect operators $\Box_\alpha$, one for each formula $\alpha$.

Let $\alpha, \beta, \varphi, \psi$ range over arbitrary formulas. The set $\mathcal{F}$ of formulas is constructed as follows:

$$\varphi =:: p \mid \neg\varphi \mid \varphi \vee \psi \mid G\varphi \mid H\varphi \mid \Box_\alpha\varphi \mid C(\alpha, \beta)$$

Other Boolean connectives are defined as usual. Let $\bot =_{df} p \wedge \neg p$, $\top =_{df} \neg\bot$, $F\varphi =_{df} \neg G\neg\varphi$, $P\varphi =_{df} \neg H\neg\varphi$, and $\diamond_\alpha\varphi =_{df} \neg\Box_\alpha\neg\varphi$.

Temporal operators have their usual meanings. $G\varphi$ says that it will always be the case that $\varphi$; $H\varphi$ says that it has always been the case that $\varphi$. Correspondingly, $F\varphi$ ($P\varphi$) states that it will be (was) the case at some future (past) point that $\varphi$. The effect operators $\Box_\alpha$ characterize the effect period of $\alpha$. Formulas of the form $\Box_\alpha\varphi$ state that $\varphi$ is true throughout the effect period of $\alpha$. Formulas of the form $C(\alpha, \beta)$ are called *causal formulas* with intended meaning that $\alpha$ causes $\beta$.

We usually use $\alpha$ and $\beta$ in causal formulas and $\varphi$ and $\psi$ outside causal formulas. This is only for a clear representation. All of $\alpha, \beta, \varphi$, and $\psi$ represent arbitrary formulas.

We choose a general syntax in which arbitrary formulas can occur in a causal formula. This may not be appropriate in some cases. Consider the formula $C(Fp, q)$, which says that an occurrence of $p$ in the future causes $q$. This causal relation depends on what will happen in the future, which contradicts the basic idea of CCT that causal relations should only depend on the past experience. We can restrict the syntax in an obvious way to exclude such cases. For the sake of simplicity, however, we keep the current one.

### *2.2. Model.*

DEFINITION 2.1. *A* (causal) model *is a tuple* $\mathfrak{M} = (T, <, \{R_\alpha \mid \alpha \in \mathcal{F}\}, V)$ *defined as follows:*

*(i)* $T$ *is a nonempty set of (time) points.*

*(ii)* $<$ *is a linear order on* $T$*, that is,* $<$ *is irreflexive* ($\forall x\neg(x < x)$)*, transitive* ($\forall xyz(x < y \wedge y < z \to x < z)$)*, and trichotomic* ($\forall xy(x < y \vee x = y \vee y < x)$).

*(iii)* $V : P \to 2^T$ *is a valuation function that assigns to each* $p \in P$ *a set of points at which* $p$ *is true.*

*(iv) Each* $R_\alpha$ *is a binary relation on* $T$ *(probably empty) such that*

*(a)* $R_\alpha tu$ *implies that* $t < u$,

*(b)* $R_\alpha tu$ *implies that* $R_\alpha tv$ *for all* $v$ *such that* $t < v < u$.

$R_\alpha(t) = \{u \mid R_\alpha tu\}$ *is called* $\alpha$*'s* effect period *at* $t$*. Any* $u \in R_\alpha(t)$ *is called an* effect point *with respect to* $\alpha$ *at* $t$.

The time structure is for a concept of psychological time. It is the time that we perceive. We may consider the structure as a history of human beings, so that causal relations being built are from the perspective of human beings, or may consider the structure as a history of some agent's personal experience, so that causal relations being built are from the perspective of the agent.

We will write $(t, u)$ for $\{s \mid t < s < u\}$. Similarly, we write $[t, u)$, $(t, u]$ and $[t, u]$ for their obvious meanings. Definition 2.1(iv-a) states that effect points are always in the future; Definition 2.1(iv-b) says that if $u$ is contiguous to $t$, so are any points in $(t, u)$. $u \in R_\alpha(t)$ means that $u$ is contiguous to $t$ with respect to $\alpha$. If $\alpha$ is true at $t$ and $\beta$ is true at $u$, this occurrence of $\beta$ is contiguous to this occurrence of $\alpha$, and the pair is considered an evidence that $\alpha$ causes $\beta$.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

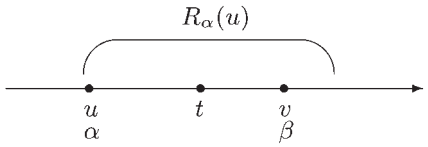



Fig. 1. The definition of data points.

***2.3. Semantics.*** Constant conjunction is a simple notion with a clear intuition. It, however, also has subtlety. Intuitively, $\alpha$ constantly conjoins with $\beta$, if, in the past, after any occurrence of $\alpha$, some $\beta$ occurred within $\alpha$'s effect period. This statement is not sound in a special case, as pictured in Figure 1.

There are three points $u, t, v$ with $u < t < v$, and $t, v \in R_\alpha(u)$. Let $t$ be the current point. Suppose $\alpha$ holds at $u$, and $\beta$ does not hold at any $s \in (u, t)$. It seems a counterexample to $\alpha$ constantly conjoining with $\beta$ at $t$, because $\alpha$ did occur at $u$, but $\beta$ did not contiguously follow from the point of view at $t$. However, it is possible that there is a future point $v$, also within the effect period $R_\alpha(u)$, at which $\beta$ holds. Thus, $u$ cannot be a counterexample. The point is that, we cannot determine at $t$ whether a $\beta$ contiguously follows the occurrence of $\alpha$ at $u$. This suggests the definition of data points.

Given a point $t$ and a formula $\alpha$, let the set of *data points*

$$D(t, \alpha) = \{u \mid u < t, \text{ and } v < t \text{ for all } v \in R_\alpha(u)\}.$$

Thus, $D(t, \alpha)$ consists of those points whose effect period has been over from the point of view of the current point.

On the other hand, even though nondata points cannot be counterexamples, they can be evidences, as shown in Figure 2. $t$ is still the current point. $u$ is not a data point with respect to $\alpha$ at $t$. Suppose that there was a $s \in (u, t)$ at which $\beta$ is true. Then the pair of occurrences of $\alpha$ (at $u$) and $\beta$ (at $s$) is a legitimate evidence of the constant conjunction, even this $\alpha$'s effect period has not been totally observed. What will happen after $t$ is irrelevant in this case, because we have already observed a $\beta$ after this $\alpha$.

DEFINITION 2.2. *$\mathfrak{M}, t \models_C \varphi$ denotes that $\varphi$ is true at point $t$ in causal model $\mathfrak{M}$, which is defined as follows:*

$\mathfrak{M}, t \models_C p$ *iff* $t \in V(p)$, *where* $p \in P$.
$\mathfrak{M}, t \models_C \neg\varphi$ *iff* $\mathfrak{M}, t \not\models_C \varphi$.
$\mathfrak{M}, t \models_C \varphi \vee \psi$ *iff* $\mathfrak{M}, t \models_C \varphi$ *or* $\mathfrak{M}, t \models_C \psi$.
$\mathfrak{M}, t \models_C G\varphi$ *iff for all* $u$ *with* $t < u$, $\mathfrak{M}, u \models_C \varphi$.
$\mathfrak{M}, t \models_C H\varphi$ *iff for all* $u$ *with* $u < t$, $\mathfrak{M}, u \models_C \varphi$.
$\mathfrak{M}, t \models_C \Box_\alpha\varphi$ *iff for all* $u$ *with* $R_\alpha tu$, $\mathfrak{M}, u \models_C \varphi$.
$\mathfrak{M}, t \models_C C(\alpha, \beta)$ *iff the following holds:*

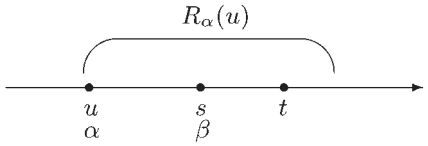



Fig. 2. A nondata point evidence.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

(i) *For each* $u \in D(t, \alpha)$ *with* $\mathfrak{M}, u \models_C \alpha$, *there is a* $v$ *such that* $R_\alpha uv$ *and* $\mathfrak{M}, v \models_C \beta$.

(ii) *There exist* $u, v$ *such that* $u < t$, $v < t$, $R_\alpha uv$, $\mathfrak{M}, u \models_C \alpha$, *and* $\mathfrak{M}, v \models_C \beta$.

(iii) *There exists a* $u < t$ *with* $\mathfrak{M}, u \not\models_C \alpha$.

(iv) *There exists a* $v < t$ *with* $\mathfrak{M}, v \not\models_C \beta$.

In the semantic clause for $C$, the first two conditions state the constant conjunction between $\alpha$ and $\beta$. Condition (i) is to the effect that there has been no *counterexample* to $C(\alpha, \beta)$. Condition (ii) says that there has been some *evidence* of $C(\alpha, \beta)$. The last two conditions exclude any invariably true formulas to be either cause or effect. We take these two as the basic conditions that distinguish causal relations from noncausal constant conjunction relations. To sum up, $\alpha$ causes $\beta$ if both of them have not been invariable true in the past, and no counterexample to it has been observed, and some evidence of it has been observed.

We sometimes write $t \models_C \varphi$ for $\mathfrak{M}, t \models_C \varphi$. A formula $\varphi$ is true in a model $\mathfrak{M}$ (denoted by $\mathfrak{M} \models_C \varphi$) if $\mathfrak{M}, t \models_C \varphi$ for all points $t$ in $\mathfrak{M}$. $\varphi$ is valid (denoted by $\models_C \varphi$) if $\mathfrak{M} \models_C \varphi$ for all models $\mathfrak{M}$. $\varphi$ is satisfiable if there is a model $\mathfrak{M}$ and a point $t$ such that $\mathfrak{M}, t \models_C \varphi$. Given a class of models $M$, $M \models_C \varphi$ if $\mathfrak{M} \models_C \varphi$ for all models $\mathfrak{M} \in M$. Notions like $t \models_C \Gamma$, where $\Gamma$ is a set of formulas, are similarly defined.

The causal operator is not a normal binary modality. It is neither "$\diamond$-type" nor "$\Box$-type." The semantic clause (i) is a $\forall$-$\exists$ statement. It is easy to check that neither the K axiom $C(\alpha \to \beta_1, \beta_2) \to (C(\alpha, \beta_1) \to C(\beta_1, \beta_2))$, nor the necessity rules $\alpha / C(\alpha, \beta)$ and $\beta / C(\alpha, \beta)$ hold.

Though the primary relations in the model are only for characterizing temporal operators and effect operators, the causal operator $C$ is not definable by the rest.

THEOREM 2.3. *$C$ is not definable by $G$, $H$, and $\{\Box_\alpha \mid \alpha \in \mathcal{F}\}$.*

*Proof.* Suppose otherwise. Let $C(p, q) =_{df} \chi$, where $\chi$ is a formula with no occurrence of $C$.

Let $\mathfrak{M}$ be a model pictured in Figure 3, which contains a points $u$ with proper past points, and after $u$, there is an integer line with their usual order.

Let the model satisfy the following:

(1) $u \models_C p$ and $0 \models_C C(p, q)$.

(2) $R_p(u) = \{i \mid i \in \mathbb{Z}, i < 2\}$.

(3) $R_\alpha(i) = \emptyset$ for all integers $i$ and all formulas $\alpha$.

(4) $i \models_C \neg q$ for all integers $i$.

(5) All integers agree on all propositional variables, that is, for all $i, j \in \mathbb{Z}$ and all $p \in P$, $i \in V(p)$ iff $j \in V(p)$.

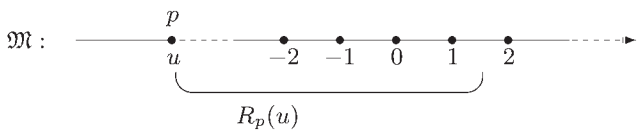



Fig. 3. The independency of $C$.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

The definition is sound because $u$ is not a $p$-data point at 0 so that it is not a counterexample to $C(p, q)$ at 0. Let the model $\mathfrak{M}'$ have the same definition as $\mathfrak{M}$ except that $R'_p(u) = \{i' \mid i' \in \mathbb{Z}, i' < 0\}$.

We have $\mathfrak{M}', 0' \models_C \neg C(p, q)$ because $u$ is now a $p$-data point at $0'$ and is a counterexample to $C(p, q)$. Then we have $\mathfrak{M}, 0 \models_C \chi$ and $\mathfrak{M}', 0' \not\models_C \chi$. It is sufficient to show that for any formula $\varphi$ with no occurrence of $C$, $\mathfrak{M}, 0 \models \varphi$ iff $\mathfrak{M}', 0' \models_C \varphi$.

Let $Z$ be the relation between the two models defined as follows:

(i) For any noninteger points $t$, $Z$ is one-one so that $tZt'$.

(ii) For integer points, $Z$ is total, that is, $iZj'$ for all integers $i$ and $j'$.

It is a routine to check that $Z$ is a bisimulation. Thus, 0 and $0'$ satisfy exact the same set of formulas with no occurrence of $C$. □

LEMMA 2.4. $\models_C \neg C(\alpha, \beta) \wedge PC(\alpha, \beta) \rightarrow G\neg C(\alpha, \beta)$.

*Proof.* Suppose $t \models_C \neg C(\alpha, \beta)$ and $t \models_C PC(\alpha, \beta)$. Then there is a $u < t$ with $u \models_C C(\alpha, \beta)$, that is, all the four conditions for $C(\alpha, \beta)$ in Definition 2.2 hold at $u$. Because $u < t$, conditions (ii)–(iv) must also hold at $t$. Because $C(\alpha, \beta)$ is false at $t$, condition (i) must fail at $t$, that is, there is a data point $v \in D(t, \alpha)$ such that $v$ is a counterexample to $C(\alpha, \beta)$ at $t$. Clearly, $v$ is also a counterexample to $C(\alpha, \beta)$ at all $s$ with $t < s$. Then $s \models_C \neg C(\alpha, \beta)$. Thus, $t \models_C G\neg C(\alpha, \beta)$. □

Lemma 2.4 states that once we lose a causal relation, we will lose it forever. This is because losing it means that we found a counterexample to the causal relation. According to the lemma, the distribution of a causal statement $C(\alpha, \beta)$ along a time line is quite ordered. There are only five cases, as shown in Figure 4:

(i) is the case that the causal relation does not hold throughout the structure. (ii) is the case that the causal relation always holds. (iii) is the case that the causal relation holds after some time point.

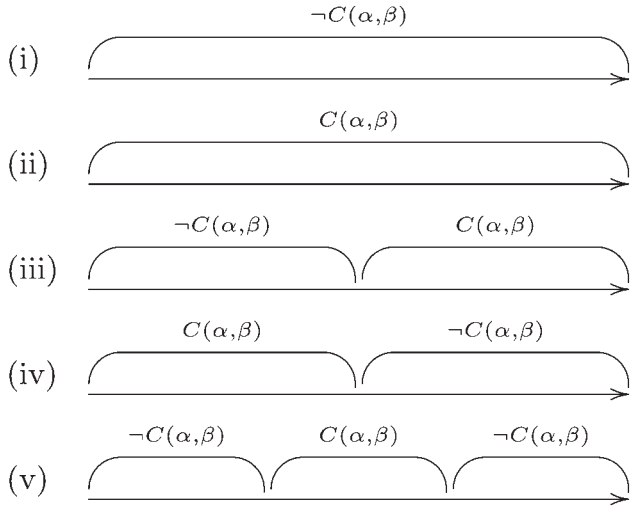



Fig. 4. The distributions of causal formulas.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

In case (ii) and case (iii), $C(\alpha, \beta)$ will always hold once it has been built. This indicates that no counterexample to the causal relation occurs anywhere. In this sense, $C(\alpha, \beta)$ can be considered as an *objective* causal relation. Though causation is an epistemic concept here, we may define the "objectivity" of causation as this *never observing a counterexample*.

(iv) and (v) are cases that the causal relation has been built, but will be found false by observing some counterexample. By Lemma 2.4, $C(\alpha, \beta)$ can never be recovered thereafter.

Case (ii) and (iv) seem unrealistic because the causal relation $C(\alpha, \beta)$ comes from nowhere. In real world, we shall start with $\neg C(\alpha, \beta)$, and build $C(\alpha, \beta)$ by collecting enough and proper data. (ii) and (iv) can be eliminated by assuming that there is a starting time point, at which no causal formula can be true according to the semantics.

The rest three cases classify causal relations into three types: Those *objective* causal relations that once was built and will always be held, as in (iii); those *false* causal relations that can never be built, as in (i); and those *delusive* causal relations that once was built, but will be removed later, as in (v).

***2.4. Axiomatization.*** In the axiomatization, for characterizing nontemporal operators $\Box_\alpha$ and $C$, we will use name formulas to distinguish time points. For this purpose, Gabbay's general irreflexivity rule (Gabbay *et al.*, 1994) suffices. Let $name(\psi) = \psi \wedge G\neg\psi$. In any transitive and trichotomic frame, if $name(\psi)$ is true at some point $t$, then it must be false at all other points. Thus, $name(\psi)$ acts as a name of $t$, and is called a *name formula*. For reference, the hybrid logic is another way of using nominals, which in its objective language contains atoms that each is interpreted as a name of a state (Areces & ten Cate, 2006).

The minimal causal logic $C$ consists of the following axioms and rules:

(*Prop*) Propositional tautologies,

(K) $G(\varphi \to \psi) \to (G\varphi \to G\psi)$, $H(\varphi \to \psi) \to (H\varphi \to H\psi)$, $\Box_\alpha(\varphi \to \psi) \to (\Box_\alpha\varphi \to \Box_\alpha\psi)$,

(Re) $\varphi \to GP\varphi$, $\varphi \to HF\varphi$,

(4) $G\varphi \to GG\varphi$,

($.3_r$) $(F\varphi \wedge F\psi) \to F(\varphi \wedge F\psi) \vee F(\varphi \wedge \psi) \vee F(\psi \wedge F\varphi)$,

($.3_l$) $(P\varphi \wedge P\psi) \to P(\varphi \wedge P\psi) \vee P(\varphi \wedge \psi) \vee P(\psi \wedge P\varphi)$,

(EP) $\Box_\alpha\varphi \wedge \diamond_\alpha name(\psi) \to G(Fname(\psi) \to \varphi)$,

(EF) $\diamond_\alpha\varphi \to F\varphi$,

(C) $name(\psi) \to [C(\alpha, \beta) \leftrightarrow P\neg\alpha \wedge P\neg\beta \wedge H(\neg \diamond_\alpha name(\psi) \wedge \alpha \to \diamond_\alpha\beta) \wedge P(\alpha \wedge \diamond_\alpha(Fname(\psi) \wedge \beta))]$,

(MP) $\dfrac{\varphi, \varphi \to \psi}{\psi}$,

(RN) $\dfrac{\varphi}{G\varphi}$, $\dfrac{\varphi}{H\varphi}$, $\dfrac{\varphi}{\Box_\alpha\varphi}$,

(IRR) $\dfrac{\langle\rangle_1(\varphi_1 \wedge \ldots \wedge \langle\rangle_m(\varphi_m \wedge name(p))\ldots) \to \varphi}{\langle\rangle_1(\varphi_1 \wedge \ldots \wedge (\langle\rangle_m\varphi_m)\ldots) \to \varphi}$,

where $p$ does not occur anywhere except in the displayed $name(p)$, and each $\langle\rangle_1 \ldots \langle\rangle_m$ is either $F$ or $P$ or $\diamond_\alpha$ for some $\alpha$.

Notions of proof ($\vdash_C \varphi$) and deduction ($\Gamma \vdash_C \varphi$) are defined as usual. The system contains the temporal logic for linear orders. The *axiom of effect period* (EP) guarantees

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

that an effect period is an uninterrupted chain. The name formula in (EP) is to identify an effect point, say $u$. The consequent of (EP) says that any point lies between the current point and $u$ is also an effect point. The *axiom of effect future* (EF) guarantees that an effect period is in the future. The *axiom of causation* (C) is a direct formalization of the semantic clause for operator $C$. Again, we need a name formula in the axiom to identify the current point in which $C(\alpha, \beta)$ is evaluated.

It is left open to us whether the irreflexivity rule add new theorems to the logic, that is, whether it is an equivalent system if we remove (IRR) from $C$. In general, little has been known on the problem that whether there are natural criteria deciding such rules' redundancy over given deductive systems (Venema, 1993).

Soundness of the system is checked as usual.

THEOREM 2.5. *(Soundness) For any formula $\varphi$, if $\vdash_C \varphi$ then $\models_C \varphi$.*

A set of formulas $\Gamma$ is ($C$-)consistent if $\Gamma \vdash_C \varphi \wedge \neg\varphi$ for no $\varphi$. A formula $\varphi$ is consistent if $\{\varphi\}$ is. $\Gamma$ is ($C$-)maximal consistent if it is consistent and any proper superset of it is inconsistent. The ($C$-)canonical model is defined as usual as a tuple $\mathfrak{M}^c = (T^c, <^c, \{R^c_\alpha \mid \alpha \in \mathcal{F}\}, V^c)$ such that $T^c$ is the set of maximal consistent sets, $t <^c u$ iff $\varphi \in u$ for all $G\varphi \in t$, $R^c_\alpha tu$ iff $\varphi \in u$ for all $\Box_\alpha \varphi \in t$, and $t \in V^c(p)$ iff $p \in t$. We say that $t$ and $u$ are connected if either $u <^c t$ or $t <^c u$. It is direct from the definition that if $name(\psi) \in t$ then $name(\psi) \notin u$ for any $u$ that connects with $t$. It follows that $name(\psi) \in t$ implies that $t$ is an irreflexive point.

For any $T \subseteq T^c$, let $\mathfrak{M}^c|T$ be the submodel of $\mathfrak{M}^c$ induced by $T$. That is, $\mathfrak{M}^c|T = (T, <, \{R_\alpha \mid \alpha \in \mathcal{F}\}, V)$, where $<=<^c \cap (T \times T)$, $R_\alpha = R^c_\alpha \cap (T \times T)$, and $V(p) = V^c(p) \cap T$ for all $p \in P$.

DEFINITION 2.6. *A subset $T \subseteq T^c$ is called* saturated *if the following holds:*

*(i) For any $t \in T$, if $F\varphi \in t$, then there is a $u \in T$ such that $t <^c u$ and $\varphi \in u$.*

*(ii) For any $t \in T$, if $P\varphi \in t$, then there is a $u \in T$ such that $u <^c t$ and $\varphi \in u$.*

*(iii) For any $t \in T$, if $\diamond_\alpha \varphi \in t$, then there is a $u \in T$ such that $R^c_\alpha tu$ and $\varphi \in u$.*

DEFINITION 2.7. *A maximal consistent set is called* witnessing *if it contains some name formula. A subset $T \subseteq T^c$ is called witnessing if every member of $T$ is witnessing. In the case that $T$ is saturated (and/or witnessing), we say that $\mathfrak{M}^c|T$ is saturated (and/or witnessing).*

LEMMA 2.8. *Let $T \subseteq T^c$ be saturated and witnessing. Then for any $t, u \in T$ with $name(\psi) \in t$ for some formula $\psi$, $u \in D(t, \alpha)$ in $\mathfrak{M}^c|T$ iff $u < t$ and $\neg \diamond_\alpha name(\psi) \in u$.*

*Proof.* Suppose $u \in D(t, \alpha)$ in $\mathfrak{M}^c|T$. By the definition of data points, $u < t$. Also, $v < t$ for all $v \in R_\alpha(u)$. Because $T$ is witnessing, $name(\psi) \notin v$ for all such $v$. Because $T$ is saturated, $\neg \diamond_\alpha name(\psi) \in u$.

Suppose $u < t$ and $\neg\diamond_\alpha name(\psi) \in u$. By axiom (EF), $R^c_\alpha \subseteq <^c$. It follows that $R_\alpha \subseteq <$. Then for all $v \in R_\alpha(u)$, $u < v$. By temporal axioms, $\mathfrak{M}^c$ is nonbranching, therefore $\mathfrak{M}^c|T$ is no-branching. Then by $u < v$ and $u < t$, we have that $v = t$, $v < t$, or $t < v$. Clearly $v \neq t$ because $name(\psi) \notin v$. We also have $t \not< v$. Suppose otherwise that $t < v$. We show that $R^c_\alpha ut$, and therefore $\diamond_\alpha name(\psi) \in u$, which contradicts the assumption. It suffices to show that for any $\Box_\alpha \varphi \in u$, $\varphi \in t$. Let $\Box_\alpha \varphi \in u$. Since $v$ is witnessing, $name(\chi) \in v$ for some $\chi$. Then $\diamond_\alpha name(\chi) \in u$. By axiom (EP), $G(Fname(\chi) \rightarrow \varphi) \in u$. Then $Fname(\chi) \rightarrow \varphi \in t$. By $t < v$, $Fname(\chi) \in t$. Then $\varphi \in t$. We conclude that $v < t$. It follows that $u \in D(t, \alpha)$. □

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

LEMMA 2.9. *If $T$ is saturated and witnessing, then for any $t \in T$ and any formula $\varphi$, $\mathfrak{M}^c|T, t \models_C \varphi$ iff $\varphi \in t$.*

*Proof.* The proof is by induction on $\varphi$. We only check the case for $C$ operator. As $T$ is witnessing, let $name(\psi) \in t$.

Suppose $\mathfrak{M}^c|T, t \models_C C(\alpha, \beta)$. Then by definition we have the followings:

(i) For each $u \in D(t, \alpha)$ in $\mathfrak{M}^c|T$ such that $\mathfrak{M}^c|T, u \models_C \alpha$, there is a $v \in R_\alpha(u)$ such that $\mathfrak{M}^c|T, v \models_C \beta$.

(ii) There exist $u, v \in T$ such that $u < t$, $v < t$, $R_\alpha uv$, $\mathfrak{M}^c|T, u \models_C \alpha$, and $\mathfrak{M}^c|T, v \models_C \beta$.

(iii) There exists a $u < t$ with $\mathfrak{M}, u \not\models_C \alpha$.

(iv) There exists a $v < t$ with $\mathfrak{M}, v \not\models_C \beta$.

By (i), Lemma 2.8, and the inductive hypothesis, for each $u$ in $T$ with $u < t$, $\neg \diamond_\alpha name(\psi) \in u$, and $\alpha \in u$, there is a $v \in R_\alpha(u)$ with $\beta \in v$. By the definition of the canonical model, for each $u$ in $T$ with $u < t$, $\neg \diamond_\alpha name(\psi) \wedge \alpha \rightarrow \diamond_\alpha \beta \in u$. Because $\mathfrak{M}^c|T$ is saturated, $H(\neg \diamond_\alpha name(\psi) \wedge \alpha \rightarrow \diamond_\alpha \beta) \in t$.

By (ii) and the inductive hypothesis, there exist $u, v \in T$ such that $u < t$, $v < t$, $R_\alpha uv$, $\alpha \in u$, and $\beta \in v$. Then $Fname(\psi) \in v$. Then $\alpha \wedge \diamond_\alpha(Fname(\psi) \wedge \beta) \in u$. Then $P(\alpha \wedge \diamond_\alpha(Fname(\psi) \wedge \beta)) \in t$.

By (iii) and (iv), $P\neg\alpha \wedge P\neg\beta \in t$. Then by axiom (C), $C(\alpha, \beta) \in t$. The other direction is similarly proved. □

LEMMA 2.10. *Let $T$ be saturated and witnessing, and $t \in T$. Let $\mathfrak{M}^c|T_t$ be the submodel of $\mathfrak{M}^c|T$ generated by $t$. Then $\mathfrak{M}^c|T_t$ is a causal model.*

*Proof.* By the temporal axioms, $\mathfrak{M}^c$ is transitive and nonbranching. Clearly $\mathfrak{M}^c|T$ is also transitive and nonbranching, and therefore $\mathfrak{M}^c|T_t$ is transitive and trichotomic. For each $u \in \mathfrak{M}^c|T_t$, because $name(\psi) \in u$ for some $\psi$, $u$ is an irreflexive point. Then $\mathfrak{M}^c|T_t$ is a linear order.

Let $<$ and $R_\alpha$ denote the respective relations in $\mathfrak{M}^c|T_t$. If $R^c_\alpha \subseteq <^c$, then $\mathfrak{M}^c|T_t$ will inherit this property, that is, $R_\alpha \subseteq <$. To show that $R^c_\alpha \subseteq <^c$, let $R^c_\alpha uv$, and $G\varphi \in u$. It suffices to show that $\varphi \in v$. By $G\varphi \in u$ and axiom (EF), $\Box_\alpha \varphi \in u$. By $R^c_\alpha uv$, $\varphi \in v$.

Suppose $R_\alpha uv$ and $u < s < v$. We have to show that $R_\alpha us$. Because $R_\alpha$ and $<$ is respective restrictions from $R^c_\alpha$ and $<^c$, $R^c_\alpha uv$ and $u <^c s <^c v$. Let $\Box_\alpha \psi \in u$. Since $v$ is witnessing, $name(\chi) \in v$ for some $\chi$. Then $\diamond_\alpha name(\chi) \in u$. By axiom (EP), $G(Fname(\chi) \rightarrow \psi) \in u$. Then $Fname(\chi) \rightarrow \psi \in s$. We have $Fname(\chi) \in s$ because $s <^c v$. Then $\psi \in s$. Then, $R^c_\alpha us$. Thus, $R_\alpha us$. □

LEMMA 2.11. *Given a consistent formula $\varphi$, there is a saturated and witnessing set $T$ of maximal consistent sets, such that $\varphi \in \Gamma$ for some $\Gamma \in T$.*

*Proof.* The proof is by a zigzag construction developed by Gabbay (Gabbay *et al.*, 1994, p. 210). The full proof is in the Appendix. □

THEOREM 2.12. *(Completeness of C) For any formula $\varphi$, if $\models_C \varphi$ then $\vdash_C \varphi$.*

*Proof.* It is sufficient to show that any consistent formula is satisfiable in a causal model. Let $\varphi$ be a consistent formula. Then by Lemma 2.11, there is a saturated and witnessing set $T$ with $\varphi \in \Gamma$ for some $\Gamma \in T$. By Lemma 2.10, $\mathfrak{M}^c|T_\Gamma$ is a causal model. Because $\mathfrak{M}^c|T$ is saturated and witnessing, $\mathfrak{M}^c|T_\Gamma$ is saturated and witnessing. By Lemma 2.9, $\mathfrak{M}^c|T_\Gamma, \Gamma \models_C \varphi$. □

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

## §3. Extensions of the minimal causal logic.

***3.1. Extensions by conditions on models.*** In this section, we consider possible extensions of the minimal causal logic for various concepts of causation. Extensions are obtained by imposing more conditions on either causal models or the semantic definition.

The study of various conditions on time structures is a routine in temporal logic. Some of them are particularly interesting for the current logic. Because of the epistemic limitation of human beings, it is reasonable to assume that (psychological) time has a starting point $(\exists x \forall y(x \leq y))$, which corresponds to the axiom $H\bot \vee PH\bot$. At the starting point, no causal relation can be built, because no data has been observed. With this condition, cases (ii) and (iv) in Figure 3 are eliminated.

The effect period of an event is modeled by a binary relation. The length of an effect period is not expressible in the general linear time. Thus, we cannot say in the model that, for example, the same event always has the same length of effect period at all points. In order to enrich our framework with this apparatus, instead of general linear time, we may work with real number time or natural number time, in which the notion of length is natural. We may assign a fixed number to each event as the length of its effect period.

Another interesting topic is to extend the logic with a notion of causal degree by counting the number of instances of evidences of causal relations. We may say that we have confidence in a causal relation with the degree 100 if there have been a hundred instances of evidences of this causal relation. In this way, we can compare reliability of causal relations by their degrees. Again, for this purpose, we have to work with a time structure in which we can count numbers, such as natural number time.

We can already say something about effect periods in general linear time. For two events $\alpha$ and $\beta$, that $\beta$'s influence lasts longer than $\alpha$'s can be modeled by $R_\alpha \subset R_\beta$. Intuitively, equivalent formulas denote the same event, so that the effect period of equivalent formulas should be the same. Formally, if $\alpha$ and $\beta$ are equivalent then $R_\alpha = R_\beta$. We may add this condition to the model. It shall be noted that this condition only makes sense when we assume a type of logical omniscience. Effect period is an epistemic notion. The effect period of an event depends on our psychological perspectives of the event. Therefore, by assuming this condition, we assume that we always know the equivalence of $\alpha$ and $\beta$ if they actually are.

The effect period of some complex formula may be relevant to the effect periods of its components. We think it is reasonable to assume that $R_{\alpha\vee\gamma} = R_\alpha \cup R_\gamma$. With this assumption, the formula

$$C(\alpha, \beta) \wedge C(\gamma, \beta) \wedge P\neg(\alpha \vee \gamma) \rightarrow C(\alpha \vee \gamma, \beta)$$

will be valid, which represents a plausible property of causation (see Section 4).

On the other hand, it seems implausible that the effect period of $\alpha \wedge \beta$ depends on effect periods of $\alpha$ and $\beta$. A massive event with a long effect period, like an earthquake, may be viewed as an agglomeration of smallish events each with a tiny effect period, like a motion of a grain of sand. There seems no way to compute an effect period of the earthquake from the effect period of the sand moving.

Following this, we should not determine the effect period of $\neg\alpha$ by the effect period of $\alpha$. Otherwise, because $\alpha \wedge \beta = \neg(\neg\alpha \vee \neg\beta)$, the effect period of $\alpha \wedge \beta$ can be reduced.

It is natural to assume that, for each occurrence of $\alpha$, we know where the influence of this occurrence ends. It requires that a nonempty effect period is always a right closed interval,

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

that is, if $R_\alpha(t) \neq \emptyset$ then $R_\alpha(t) = (t, s]$ for some $s$. With this condition, the following plausible formula becomes valid:

$$\neg C(\alpha, \beta) \wedge PC(\alpha, \beta) \rightarrow P(C(\alpha, \beta) \wedge G\neg C(\alpha, \beta)).$$

The formula states that if an existing causal relation is removed, then we know where it is removed.

***3.2. Extensions by conditions on semantics.*** In the minimal causal logic, besides the basic conditions for constant conjunction, the extra condition is that invariably true formulas are neither causes nor effects. More conditions may be added to distinguish causal relations from noncausal constant conjunction relations. The question is, which ones? There does not exist a single *right* answer to the question. We present two possible candidates.

The first is motivated by the manipulability theory of causation (Woodward, 2003). According to the theory, a cause $\alpha$ is the one that, when its status is changed, will change the status of its effects. We may require that occurrences of $\alpha$ *did* change the status of $\beta$. For, even if $\alpha \wedge \beta$ was true at some point $u$, and $\beta$ followed at some contiguous point $\upsilon$, it may not be enough to say that this is an evidence that $\alpha$ causes $\beta$. Because $\beta$ already held at $u$, its occurrence at $\upsilon$ may be caused by nothing: it is just because $\beta$'s status did not change from $u$ to $\upsilon$. Thus, we may require that a positive evidence to $C(\alpha, \beta)$ is a point at which $\alpha \wedge \neg\beta$ held, and $\beta$ contiguously followed. In this case $\alpha$ does make a difference on the truth value of $\beta$. This condition is along the line of a weak manipulability theory called interventionism (Pearl, 2000), which uses manipulation as a feature of causation without claiming that it is more fundamental than causation.

Formally, let the logic $C'$ be obtained from the minimal causal logic by changing the item (ii). in Definition 2.2 to the following:

(ii') There exist $u, \upsilon$ such that $u < t, \upsilon < t, R_\alpha u\upsilon, \mathfrak{M}, u \models_C \alpha \wedge \neg\beta$, and $\mathfrak{M}, \upsilon \models_C \beta$.

A complete system for $C'$ is obtained from system $C$ by replacing the axiom (C) with the following one:

(C') $$name(\psi) \rightarrow [C(\alpha, \beta) \leftrightarrow P\neg\alpha \wedge P\neg\beta \wedge H(\neg \diamond_\alpha name(\psi) \wedge \alpha \rightarrow \diamond_\alpha\beta) \wedge P(\alpha \wedge \neg\beta \wedge \diamond_\alpha(Fname(\psi) \wedge \beta))].$$

In the minimal causal logic, for a causal formula $C(\alpha, \beta)$ to be true, some (*positive*) evidence and no counterexample should present. For another candidate of extensions of the minimal logic, we may also require some *negative* evidence for a causal formula to be true. That is, instances that $\alpha$ did not occur, and $\beta$ did not contiguously follow. It is in this case more likely that $\alpha$ causes $\beta$.

Formally, we add the following item to the semantic clause of $C$ in Definition 2.2:

(v) There is a $u \in D(t, \neg\alpha)$ such that $R_{\neg\alpha}(u) \neq \emptyset$, $\mathfrak{M}, u \models_C \neg\alpha$, and for all $\upsilon \in R_{\neg\alpha}(u)$, $\mathfrak{M}, \upsilon \models_C \neg\beta$.

In a word, this notion of causation needs three conditions: some positive evidence, some negative evidence, and no counterexample. We call this logic $C''$. For this semantics, a complete system $C''$ is obtained from the system $C$ by replacing the axiom (C) with the following one:

(C") $$name(\psi) \rightarrow [C(\alpha, \beta) \leftrightarrow P\neg\alpha \wedge P\neg\beta \wedge H(\neg \diamond_\alpha name(\psi) \wedge \alpha \rightarrow \diamond_\alpha\beta) \wedge P(\alpha \wedge \diamond_\alpha(Fname(\psi) \wedge \beta)) \wedge P(\neg \diamond_{\neg\alpha} name(\psi) \wedge \diamond_{\neg\alpha} T \wedge \neg\alpha \wedge \Box_\alpha\neg\beta)].$$

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

Logics $C'$ and $C''$ can be combined in an obvious way. Let the logic $C'''$ be obtained from the minimal causal logic by changing the item (ii) in Definition 2.2 to (ii'), and adding the following:

(v') There is a $u \in D(t, \neg\alpha)$ such that $R_{\neg\alpha}(u) \neq \emptyset$, $\mathfrak{M}, u \models_C \neg\alpha \wedge \neg\beta$, and for all $v \in R_{\neg\alpha}(u)$, $\mathfrak{M}, v \models_C \neg\beta$.

A complete system $C'''$ with respect to this semantics can be obtained by replacing the axiom (C) with (C"'):

(C"') $$name(\psi) \rightarrow [C(\alpha, \beta) \leftrightarrow P\neg\alpha \wedge P\neg\beta \wedge H(\neg \diamond_\alpha name(\psi) \wedge \alpha \rightarrow \diamond_\alpha\beta) \wedge P(\alpha \wedge \neg\beta \wedge \diamond_\alpha(Fname(\psi) \wedge \beta)) \wedge P(\neg \diamond_{\neg\alpha} name(\psi) \wedge \diamond_{\neg\alpha} T \wedge \neg\alpha \wedge \neg\beta \wedge \Box_\alpha\neg\beta)].$$

In the minimal causal logic, it is possible that (1) for some point $t$, $t \models_C C(\alpha, \beta) \wedge C(\neg\alpha, \beta)$, and (2) for some point $u$, $u \models_C C(\alpha, \beta) \wedge C(\alpha, \neg\beta)$. They are unrealistic statements about causation. Contradictory events cannot both be causes of the same effect, and also cannot both be effects of the same cause. It can be verified that (1) is eliminated in $C''$, whereas (2) holds in all of $C'$, $C''$, and $C'''$.

In practice, when we state a causal relation, like "short-circuit causes catching a fire," it usually assumes that "catching fire" occurs much less than "not catching fire." The latter is a routine that almost occurs at all situations. For such objects we do not look for a cause. This suggests a possible solution to (2) by distinguishing accidental events from the routines. We leave it for future research.

***3.3. Transitivity.*** The transitivity of causation coincides with commonsense notions of causation, and fits with some explanatory practices. On the other hand, there have been quite a few counterexamples to the transitivity of causation (see Hall, 2000, for counterexamples and Lewis, 2004, for defensive replies).

Causation in our minimal logic is not transitive: $C(\alpha, \beta) \wedge C(\beta, \gamma) \rightarrow C(\alpha, \gamma)$ is not valid. A transitive notion of causation can be obtained in an obvious way, as what David Lewis did in his counterfactual theory: It is simply by taking the transitive closure of the causal relations.

A *causal chain* from $\alpha$ to $\beta$ at a time point $t$ is a finite sequence $\alpha, \gamma_1, \ldots, \gamma_n, \beta$ such that $t \models_C C(\alpha, \gamma_1), t \models_C C(\gamma_1, \gamma_2), \ldots, t \models_C C(\gamma_n, \beta)$. Let $C^*$ be the transitive causal operator. Then,

- $t \models_C C^*(\alpha, \beta)$ if and only if there is a causal chain from $\alpha$ to $\beta$ at $t$.

THEOREM 3.1. *$C^*$ is not definable in the minimal causal logic.*

*Proof.* Suppose otherwise. Let $C^*(p, q) =_{df} \chi$, where $\chi$ is a formula in the minimal causal logic. Construct a model $\mathfrak{M}_1$ with no starting point such that at a point $t$, the followings hold:

(1) $\mathfrak{M}_1, t \not\models_C C^*(p, q)$.
(2) $p$ and $q$ are not invariably true in the past of $t$.
(3) There are some points $u, v$ in the past of $t$ such that $\mathfrak{M}_1, u \models_C p$ and $\mathfrak{M}_1, v \models_C q$.
(4) $u$ in (3) is a data point in $D(t, p)$.
(5) $R_r = \emptyset$ for any $r$ that does not occur in $\chi$.
(6) $R_p(s) \neq \emptyset$ for all points $s$.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

Let $r$ be any propositional variable that does not occur in $\chi$. Let $V_1$ be the valuation function of $\mathfrak{M}_1$. A model $\mathfrak{M}_2$ is the same as $\mathfrak{M}_1$, except that we change the definitions of $R_r$ and the valuation function on $r$ as follows:

(1') $r$ is false at some point in the past of $t$.

(2') For each data point $u \in D(t, p)$ with $\mathfrak{M}_2, u \models_C p$, let $\mathfrak{M}_2, v \models_C r$ for some $v \in R_p(u)$. This is sound because $R_p(u)$ is not empty.

(3') Pick up a single point $v$ in the past of $t$ with $\mathfrak{M}_2, v \models_C q$, pick up a $s$ in the past of $v$, and let $R_r sv$ and $\mathfrak{M}_2, s \models_C r$. This is sound because the model has no starting point.

By (2), (4), and (2'), $\mathfrak{M}_2, t \models_C C(p, r)$. By (2), (1'), and (3'), $\mathfrak{M}_2, t \models_C C(r, q)$. Then $\mathfrak{M}_2, t \models_C C^*(\alpha, \beta)$. We have that $\mathfrak{M}_1, t \not\models_C \chi$ and $\mathfrak{M}_2, t \models_C \chi$. However, because $r$ does not occur in $\chi$, we also have $\mathfrak{M}_1, t \models_C \chi$ if and only if $\mathfrak{M}_2, t \models_C \chi$. □

The point in the proof is that, unlike in the minimal causal logic, the truth value of $C^*(p, q)$ depends on values of propositional variables other than $p$ and $q$. Similar arguments can show that $C^*$ is not definable in the extended causal logics $C'$, $C''$, and $C'''$.

***3.4. Inus condition.*** Mackie (1965) has developed a sophisticated theory of causation along the line of CCT. The so-called inus condition is better illustrated by his example that a short-circuit causes a house's catching fire. First, the short-circuit is not a sufficient condition for the fire. It is a number of conditions, including the short-circuit, the presence of inflammable materials, the absence of a suitably placed sprinkler, and so on, that constitute a sufficient condition for the house's catching fire. Second, though the complex condition is sufficient, it is not necessary because the fire could have started in other ways. Therefore, the short-circuit is an *inus* condition in the meaning that it is an insufficient but nonredundant part of a condition which is itself unnecessary but sufficient for the fire.

Causes in this paper are not inus conditions. When $C(\alpha, \beta)$ holds, $\alpha$ is a sufficient condition for $\beta$, in the sense that $\alpha$ itself can bring about $\beta$. $\alpha$ may not be a necessary condition for $\beta$, because it is possible that $C(\gamma, \beta)$ holds for some other cause $\gamma$. Thus, $\alpha$ is an *us* condition (unnecessary but sufficient condition) for $\beta$. It is possible to define an inus condition in the current framework. For this purpose, it suffices to define the nonredundant parts of a cause $\alpha$.

Let $CNF(\alpha)$ be a conjunctive normal form of $\alpha$ such that $CNF(\alpha)$ contains only symbols occurring in $\alpha$, and there is no repeated conjunct. The restriction is to guarantee that $CNF(\alpha)$ contains no redundant part. Let $\gamma_i \in CNF(\alpha)$ denote that $\gamma_i$ is a conjunct in $CNF(\alpha)$. Then we say that $\gamma$ is a *nonredundant part* of $\alpha$ if there are $\gamma_1, \ldots, \gamma_n \in CNF(\alpha)$ such that $\gamma = \gamma_1 \wedge \ldots \wedge \gamma_n$. An inus type causal operator $C^i$ can be defined as follows:

- $t \models_C C^i(\gamma, \beta)$ iff $\gamma$ is a nonredundant part of some $\alpha$ such that $t \models_C C(\alpha, \beta)$.

We have that

THEOREM 3.2. *$C^i$ is not definable in the minimal causal logic.*

*Proof.* The proof is similar to the proof of Theorem 3.1, by showing that the truth value of $C^i(p, q)$ is not totally determined by the truth values of $p$ and $q$. □

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

**§4. Related work.** Branching time is a time structure where the past is always linear, while the future may be branching. von Kutschera (1993) and Xu (1997) have defined a notion of causation in branching time. Though the frameworks in the two proposals are different in technical details, they share the basic intuition. Their basic idea is that an event $e$ is a cause of an event $e'$ if the occurrence of $e$ guarantees the occurrence of $e'$. The focus is on defining "guarantee" by a type of historical necessity: Roughly speaking, it means that $e$'s occurrence in any history makes sure an occurrence of $e'$ in that history. This understanding of causation is apparently different from ours based on CCT, since CCT does not concern what will happen in the future, whether it is deterministic or indeterministic. Causation is totally determined by the (linear) past. Thus, linear time, instead of branching time, is sufficient for characterizing CCT. Another important difference is that events are primitive in our logic, whereas they define events from the underlying time structure. It shall be noted that though the famous STIT theory (Belnap & Perloff, 1988) has a branching time structure, using the notion of undivided histories, it concerns the abilities of agents rather than causation.

Reasoning about actions and changes is one of the oldest topics in logical AI, and is still an active research area. A notion of causation is normally required in the formalisms of actions, because their effects are what we concern about actions. For an incomplete list, there are Dynamic Logic, Situation Calculus, Event Calculus, Fluent Calculus, Features and Fluents, etc. Though most work in these areas take causation as an implicit notion behind their formal frameworks, some do focus on properties of causation itself. The logic in Lin (1996) remains in the framework of classical logic. It introduces a special predicate "*Caused*" which can only be applied to atoms. McCain & Turner (1997) defines a notion that a formula is a consequence of a given causal theory, which is a set of causal laws in the form of $\varphi \Rightarrow \psi$. Thielscher (1997) presents an interesting method of deriving causal rules automatically from given influence information. The simplest example is, given a constraint $p \vee q$ and the influence information $(p, q) \in I$, which means that $p$ might possibly influence $q$, we can generate a causal law "$\neg p$ causes $q$." Schwind (1999) gives several criteria of causation by which existing formalisms are examined. Giordano *et al*. (2000) defines a dynamic logic for actions and causality in which causality is expressed by a unary operator read as "caused." Also with a dynamic logic, Zhang & Foo (2001) proposes to treat direct effects and propagated effects of actions in a unified framework by using propositions as modalities. Pearl (2000) and Halpern & Pearl (2005) define causal relations in structural equation models based on the counterfactual theory of causation. They concern more about model theory properties than a logical theory. The list can be longer without being exhausted.

The focus of these works in AI is not to defend some notion of causation in philosophy. It is to invent formalisms, with as least cost as possible, to solve fundamental problems of AI in actions. The most concerned problem are the *frame problem* (the persistency of facts after an action has been applied), the *ramification problem* (the propagated effects after an action has been applied), and the *qualification problem* (all the necessary conditions before an action can be applied). Not surprisingly, they admit some implausible statements about causation. For example, Lin (1996), McCain & Turner (1997), and Giordano *et al*. (2000) admit monotonicity of causal relations, that is, from $\varphi$ causing $\psi$ infer that $\varphi \wedge \chi$ causes $\psi$.

David Lewis's counterfactual theory is the most popular philosophical theory of causation in decades (Lewis, 1973a). Though Lewis's analysis of causation is impressive in its philosophical subtlety, the logical theory of counterfactuals, the so-called conditional logic (Lewis, 1973b, chapter 6), can hardly be considered as a causal logic. In the logic,

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

the formula $\alpha > \beta$ is to state that $\beta$ is counterfactually depended on $\alpha$. If we read $\alpha > \beta$ as that $\alpha$ causes $\beta$, some axioms of the logic would be counterintuitive, such as $\alpha > \alpha$, $\alpha \wedge \beta \rightarrow (\alpha > \beta)$ and $(\alpha > \beta) \rightarrow (\alpha \rightarrow \beta)$.

As a remedy, Giordano & Schwind (2004) defined a causal logic (called AC) by choosing those axioms of Lewis's logic that represent wanted properties of causation, and omitting those unwanted axioms, including the three we mentioned. AC has a similarity with our logic, but also has fundamental differences. AC admits that a tautology is caused by everything, that is, whenever $\beta$ is a tautology, $\vdash \alpha > \beta$. They want to keep this formula because it is derivable from a basic inference rule of conditional logic which they want to keep: If $\vdash \beta \rightarrow \gamma$ then $\vdash (\alpha > \beta) \rightarrow (\alpha > \gamma)$. The rule in general is not valid in our logic. Instead, it can be checked that a restrictive version of the rule is valid, that is, if $\vdash_C \beta \rightarrow \gamma$ then $\vdash_C C(\alpha, \beta) \wedge P\neg\gamma \rightarrow C(\alpha, \gamma)$. This valid rule in our logic, we think, is more suitable as a rule of a causal logic than the general rule in conditional logic.

AC has the axiom $(\alpha > \beta) \wedge (\gamma > \beta) \rightarrow (\alpha \vee \gamma > \beta)$, which is not valid in the minimal causal logic. The axiom seems right about causation. We have shown in Section 3.1 that if we assume $R_{\alpha\vee\gamma} = R_\alpha \cup R_\gamma$, then a restrictive version of the axiom will be valid in our logic, that is, $C(\alpha, \beta) \wedge C(\gamma, \beta) \wedge P\neg(\alpha \vee \gamma) \rightarrow C(\alpha \vee \gamma, \beta)$. The difference is that we state in the premise that $\alpha \vee \gamma$ is not invariably true in the past. Again, we think it is a more suitable axiom for the interaction between causal relation and disjunction.

As for the interaction between causal relation and conjunction, we think in general there should be no connection between the two, because the effect period of a conjunction should not depends on its conjuncts. AC abandons the valid formula in conditional logic, $\alpha > \beta \rightarrow (\alpha \wedge \gamma) > \beta$. Nevertheless, it introduces a weaken formula: $\neg(\alpha > \neg\gamma) \wedge (\alpha > \beta) \rightarrow ((\alpha \wedge \gamma) > \beta)$. They argue for the formula for its usefulness in describing interactions among actions and facts. However, it still suffers from those counterexamples to $\alpha > \beta \rightarrow (\alpha \wedge \gamma) > \beta$: $rain$ causes $wet$ but $rain \wedge umbrella$ does not.

AC admits an axiom $(\gamma > \alpha) \wedge (\gamma > (\alpha > \beta)) \rightarrow (\gamma > \beta)$, which is not valid in the minimal causal logic. Its invalidity in the minimal causal logic is not a surprise, because what we formalize in the minimal causal logic are causal relations directly from our observations. We can extend the logic to deal with such indirect causal knowledge. The logic with the transitive causal operator $C^*$ in Section 3.3 is an example of such extensions. The formula $C(\gamma, \alpha) \wedge C(\gamma, C(\alpha, \beta)) \rightarrow C(\gamma, \beta)$ states an ability of us to deduce indirect causal relations from the existing ones. We can use it to extend our logic in a similar way as we did in Section 3.3.

**§5. Conclusion.** We give a formal framework for causation based on constant conjunction theory. Causal relation is expressed as a binary operator $C$. We investigate properties of $C$ with the help of a linear temporal logic. Causal statements are evaluated in a linear time, where their truth values are determined by what has been in the past. We present the minimal causal logic, which is mainly a direct formalization of the concept of constant conjunction by Hume. We discuss its extensions by assuming more conditions on models and semantics. There remain a lot to be investigated, partially because the concept of causation is itself bewildering. It is open to arguments that which conditions should be applied, so that the concept of causation being formalized can satisfy, loosely speaking, some philosophical standard.

The degree of causation is one of the important topics that is not explored in the paper. Such degree theory will be different from the probabilistic theory of causation. A probabilistic cause $\alpha$ is such that $\alpha$'s occurrence increases the probability of the effect $\beta$. The

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

expected degree theory for the current framework would mainly concern the reliability of causal relations: If we have observed more instances of evidences to a causal relation, we will have more confidence in it.

Proper extensions of the logics in this paper can serve for an agent and multiagent theory. For this, we may look upon a linear time model as an agent' perspective of what have happened along time. An agent builds or removes a causal relation according to her experiences as instances of evidences or counterexamples being observed. Things could be more interesting in the case of multiagent systems. We may let each agent has her own model, which is a linear time model represents her past experience. Thus different agents may possess different causal knowledge. Or, we may let agents only have incomplete observations of what have really happened. One way to do this is to assign a subset of the set of the past points to each agent as her observations.

**§6. Acknowledgments.** This research is supported by A Foundation for the Author of National Excellent Doctoral Dissertation of PR China (FANEDD 2007B01) and Guangdong 12th Five-Year Planning Fund of Philosophy and Social Science, and partially supported by the MOE Key Project of Philosophy and Social Sciences (No. 10JZD0006).

**Appendix A: Proof of Lemma 2.11.** The proof method is due to Gabbay *et al*. (1994). We adapt it for our framework and adopt the notational convention in Blackburn *et al*. (2001) to make the proof more readable.

A tree $(N, A)$ is defined as usual such that $N$ is the set of nodes, and $A$ is the set of arcs. A path in $(N, A)$ is a sequence of nodes $(t_1, \ldots, t_m)$ such that $(t_i, t_{i+1})$ is an arc for $1 \leq i < m$.

DEFINITION A.1. *A* network *is a tuple $\mathcal{N} = (N, A, d, f)$ such that $(N, A)$ is a finite tree; $d$ labels each arc $(u, v)$ of the tree with a member in $\{F, P\} \cup \{\diamond_\alpha \mid \alpha \in \mathcal{F}\}$; $f$ labels each node of the tree with a finite, nonempty, and consistent set of formulas.*

*A network $\mathcal{N}' = (N', A', d', f')$ is an* extension *of a network $\mathcal{N} = (N, A, d, f)$ if $N \subseteq N'$, $A = A' \cap N \times N$, $d(u, v) = d'(u, v)$ for any $u, v \in N$, and $f(u) \subseteq f'(u)$ for any $u \in N$.*

*Let $\lambda(u) = \bigwedge f(u)$. Let $S(u)$ be the set of successors of $u$.*

*A network $\mathcal{N}$ is* coherent *if for any nonleaf node $u$ and any $v \in S(u)$, $d(u, v)\lambda(v) \in f(u)$.*

DEFINITION A.2. *Let $\mathcal{N} = (N, A, d, f)$ be a network, $u$ a node in the network and $\varphi$ a formula. Let $t$ be the root of the network. Let $(u_1, \ldots, u_m)$ where $u_1 = t$ and $u_m = u$ be the path from $t$ to $u$. Let $\mathcal{N}_{\varphi,u} = (N, A, d, f_{\varphi,u})$ be defined as follows:*

*(1) For any node $v$ that does not occur in the path $(u_1, \ldots, u_m)$, $f_{\varphi,u}(v) = f(v)$.*

*(2) $f_{\varphi,u}(u) = f(u) \cup \{\varphi\}$.*

*(3) For each $1 \leq n < m$, $f_{\varphi,u}(u_n)$ is inductively defined as that $f_{\varphi,u}(u_n) = f(u_n) \cup \{d(u_n, u_{n+1})\lambda_{\varphi,u}(u_{n+1})\}$, where $\lambda_{\varphi,u}(u_{n+1}) = \bigwedge f_{\varphi,u}(u_{n+1})$.*

$\mathcal{N}_{\varphi,u}$ is a structure obtained from $\mathcal{N}$ by adding the formula $\varphi$ to the label of $u$ and making corresponding changes to other nodes, so that if $\mathcal{N}$ is coherent, so is $\mathcal{N}_{\varphi,u}$. In general, $\mathcal{N}_{\varphi,u}$ may not be a network. The label of some node in $\mathcal{N}_{\varphi,u}$ may be inconsistent set of formulas.

LEMMA A.3. *Given a coherent network $\mathcal{N} = (N, A, d, f)$, a node $u$ in the network and a formula $\varphi$, either $\mathcal{N}_{\varphi,u}$ or $\mathcal{N}_{\neg\varphi,u}$ is a coherent network.*

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

*Proof.* Suppose otherwise that both $\mathcal{N}_{\varphi,u}$ and $\mathcal{N}_{\neg\varphi,u}$ are not coherent networks. Let $t$ be the root of the network. Let $(u_1, \ldots, u_m)$ be the path from $t$ to $u$. The coherency of both $\mathcal{N}_{\varphi,u}$ and $\mathcal{N}_{\neg\varphi,u}$ is direct by the definition. Then both of them are not networks. There must exist a $i$, $1 \leq i \leq m$, such that $f_{\varphi,u}(u_i)$ is inconsistent, and must exist a $j$, $1 \leq j \leq m$ such that $f_{\neg\varphi,u}(u_j)$ is inconsistent. Note that in normal modal logic, if a formula $\psi$ is inconsistent, so is $\diamond\psi$. Because $F$, $P$, and $\diamond_\alpha$ are standard diamonds, for all $k \leq i$ and all $l \leq j$, $f_{\varphi,u}(u_k)$ and $f_{\neg\varphi,u}(u_l)$ are inconsistent. Then there is a $n$ such that both $f_{\varphi,u}(u_n)$ and $f_{\neg\varphi,u}(u_n)$ are inconsistent. For simplicity, let $\diamond_i$ stand for $d(u_i, u_{i+1})$. Then

$$\vdash_C \lambda(u_n) \rightarrow \neg \diamond_n (\lambda(u_{n+1}) \wedge \diamond_{n+1}(\ldots \wedge \diamond_{m-1}(\lambda(u_m) \wedge \varphi)\ldots). \tag{A1}$$

$$\vdash_C \lambda(u_n) \rightarrow \neg \diamond_n (\lambda(u_{n+1}) \wedge \diamond_{n+1}(\ldots \wedge \diamond_{m-1}(\lambda(u_m) \wedge \neg\varphi)\ldots). \tag{A2}$$

By A1 and A2, we will have

$$\vdash_C \lambda(u_n) \rightarrow \neg \diamond_n (\lambda(u_{n+1}) \wedge \diamond_{n+1}(\ldots \rightarrow \diamond_{m-1}\lambda(u_m)\ldots). \tag{A3}$$

By the coherency of $\mathcal{N}$, we will have

$$\vdash_C \lambda(u_n) \rightarrow \diamond_n(\lambda(u_{n+1}) \wedge \diamond_{n+1}(\ldots \rightarrow \diamond_{m-1}\lambda(u_m)\ldots). \tag{A4}$$

It follows that $\lambda(u_n)$ is inconsistent and $\mathcal{N}$ is not a network. □

DEFINITION A.4. *In a network $\mathcal{N}$, a D1 defect is a pair $(u, F\varphi)$ such that $F\varphi \in f(u)$ and there is no $v \in S(u)$ with $d(u, v) = F$ and $\varphi \in f(v)$ ; a D2 defect is a pair $(u, P\varphi)$ such that $P\varphi \in f(u)$ and there is no $v \in S(u)$ with $d(u, v) = P$ and $\varphi \in f(v)$; a D3 defect is a pair $(u, \diamond_\alpha\varphi)$ such that $\diamond_\alpha\varphi \in f(u)$ and there is no $v \in S(u)$ with $d(u, v) = \diamond_\alpha$ and $\varphi \in f(v)$; a D4 defect is a pair $(u, \varphi)$ such that neither $\varphi \in f(u)$ nor $\neg\varphi \in f(u)$; a D5 defect is a node $u$ such that $name(\psi) \notin u$ for any formula $\psi$.*

LEMMA A.5. *For any defect of a coherent network $\mathcal{N}$, there is a coherent extension of $\mathcal{N}$ that lacks the defect.*

*Proof.* Let $\mathcal{N} = (N, A, d, f)$. We show each type of defects is repairable.

Suppose that $(u, F\varphi)$ is a D1 defect. Let $v$ be an arbitrary node that is not in $N$. Let $\mathcal{N}' = (N', A', d', f')$ be such that $N' = N \cup \{v\}$, $A' = A \cup \{(u, v)\}$, $d' = d \cup \{(u, v), F\}$, and $f' = f \cup \{v, \{\varphi\}\}$. $\varphi$ is consistent because $F\varphi$ is. Then, $\mathcal{N}'$ is a network. Clearly, $\mathcal{N}'$ is an extension of $\mathcal{N}$ in which the D1 defect has been removed. $\mathcal{N}'$ is coherent, because $\mathcal{N}$ is coherent and $F\varphi$ is already in $u$. D2 and D3 defects can be removed in a similar way.

Suppose $(u, \varphi)$ is a D4 defect. By Lemma A.3, either $\mathcal{N}_{\varphi,u}$ or $\mathcal{N}_{\neg\varphi,u}$ is a coherent network. That coherent network is an extension of $\mathcal{N}$ that lacks the defect.

Suppose $u$ is a D5 defect. Let $t$ be the root of the network. Let $(u_1, \ldots, u_m)$ be the path from $t$ to $u$. Let $p$ be a propositional variable that does not occur anywhere in $\mathcal{N}$. Clearly, $\mathcal{N}_{name(p),u}$ lacks the D5 defect. We prove that $\mathcal{N}_{name(p),u}$ is a coherent network. Coherence is direct by definition. It suffices to show that $f_{name(p),u}(u_n)$ is consistent for every $1 \leq n \leq m$. Suppose otherwise that $f_{name(p),u}(u_n)$ is inconsistent for some $n$. Then $\vdash_C \neg\lambda_{name(p),u}(u_n)$, that is,

$$\vdash_C \neg(\lambda(u_n) \wedge d(u_n, u_{n+1})(\lambda(u_{n+1}) \wedge \ldots \wedge d(u_{m-1}, u_m)(\lambda(u_m) \wedge name(p))\ldots).$$

Then, by the IRR rule,

$$\vdash_C \lambda(u_n) \rightarrow \neg d(u_n, u_{n+1})(\lambda(u_{n+1}) \wedge \ldots \wedge d(u_{m-1}, u_m)(\lambda(u_m))\ldots).$$

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

However, because $\mathcal{N}$ is coherent, we have

$$\vdash_C \lambda(u_n) \rightarrow d(u_n, u_{n+1})(\lambda(u_{n+1}) \wedge \ldots \wedge d(u_{m-1}, u_m)(\lambda(u_m)) \ldots).$$

We obtain that $\lambda(u_n)$ is not consistent, and therefore $\mathcal{N}$ is not a network. □

DEFINITION A.6. *Given a consistent formula $\varphi$, we define a sequence of countably infinite networks $\mathcal{N}_0, \mathcal{N}_1, \ldots$ such that the following holds:*

*(i) Each network in the sequence is coherent.*
*(ii) $\mathcal{N}_0$ consists of a single node $t$ with $f(t) = \{\varphi\}$.*
*(iii) For any $i < j$, $\mathcal{N}_j$ is an extension of $\mathcal{N}_i$.*
*(iv) For any $\mathcal{N}_i$ and any defect in $\mathcal{N}_i$, there is a $j > i$ such that $\mathcal{N}_j$ lacks this defect.*

The definition is sound by Lemma A.5 and the fact that there are only countably infinite defects occurring in the sequence: The number of defects is upper bounded by the sum of $|\{F, P\} \times \mathcal{F} \times \omega|$ (for D1 and D2 defects), $|\mathcal{F} \times \mathcal{F} \times \omega|$ (for D3 defects), $|\mathcal{F} \times \omega|$ (for D4 defects) and $|\omega|$ (for D5 defects). Let $N = \bigcup_{i \in \omega} N_i$. Let $f(u) = \bigcup_{i \in \omega} f_i(u)$ for each $u \in N$.

LEMMA A.7. *For each $u \in N$, $f(u)$ is a maximal consistent set and is witnessing.*

*Proof.* If a D5 defect $u$ is in some $\mathcal{N}_i$, then there is an extension $\mathcal{N}_j$ that lacks the defect, that is, $name(\psi) \in f_j(u)$ for some $\psi$. Then $name(\psi) \in f(u)$ and $u$ is witnessing. Similarly, either $\psi$ or $\neg\psi$ is in $f(u)$ for all formulas $\psi$.

Suppose $f(u)$ is inconsistent. Then there must be a $i$ such that $f_i(u)$ is inconsistent. Then $\mathcal{N}_i$ is not a network. □

LEMMA A.8. *Let $T = \{f(u) | u \in N\}$. Then $T$ is saturated and witnessing.*

Proof. It is direct by Lemma A.7 that $T$ is witnessing. For saturatedness, let $u$ be any member in $N$ such that $F\psi \in f(u)$ . Then $F\psi \in f_i(u)$ for some $i$. By Lemma A.5, there is a $j$ such that $\psi \in f_j(v)$ and $d_j(u, v) = F$. Then $\psi \in f(v)$. It remains to show that $f(u) <^c f(v)$. Suppose otherwise. Then there is formula $\chi$ such that $\chi \in f(v)$ and $\neg F\chi \in f(u)$. Then there is a $k$ such that $\chi \in f_k(v)$ and $\neg F\chi \in f_k(u)$. By the coherency of $\mathcal{N}_k$, $F\chi \in f_k(u)$. It follows that $\mathcal{N}_k$ is not a network.

The cases for $P$ and $\diamond_\alpha$ are similarly proved. □

## BIBLIOGRAPHY

Areces, C., & ten Cate, B. (2006). Hybrid logics. In Blackburn, P., Wolter, F., and van Benthem, J., editors. *Handbook of Modal Logic*. Amsterdam, The Netherlands: Elsevier, pp. 821–868.

Baumgartner, M. (2008). Regularity theories reassessed. *Philosophia*, **36**(3), 327–354.

Baumgartner, M. (2009). Uncovering deterministic causal structures: A boolean approach. *Synthese*, **170**(1), 71–96.

Beebee, H. (2006). Does anything hold the universe together? *Synthese*, **149**(3), 509–533.

Belnap, N., & Perloff, M. (1988). Seeing to it that: a canonical form for agentives. *Theoria*, **54**(3), 175–199.

Black, M. (1956). Why cannot an effect precede its cause? *Analysis*, **16**(3), 49–58.

Blackburn, P., De Rijke, M., & Venema, Y. (2001). *Modal Logic*. Cambridge, UK: Cambridge University Press.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

Dummett, A. E., & Flew, A. (1954). Symposium: Can an effect precede its cause? *Proceedings of the Aristotelian Society, Supplementary Volumes*, **28**, 27–62.

Gabbay, D. M., Hodkinson, I., & Reynolds, M. (1994). *Temporal Logic: Mathematical Foundations and Computational Aspects*, Vol. 1. New York: Oxford University Press.

Giordano, L., Martelli, A., & Schwind, C. (2000). Ramification and causality in a modal action logic. *Journal of Logic and Computation*, **10**(5), 625–662.

Giordano, L., & Schwind, C. (2004). Conditional logic of actions and causation. *Artificial Intelligence*, **157**(1–2), 239–279.

Graßhoff, G., & May, M. (2001). Causal regularities. In Spohn, W., Ledwig, M., and Esfeld, M., editors. *Current Issues in Causation*. Paderborn, Germany: Mentis, pp. 85–114.

Hall, N. (2000). Causation and the price of transitivity. *The Journal of Philosophy*, **97**(4), 198–222.

Hall, N. (2004). Two concepts of causation. In Collins, J., Hall, N., and Paul, L. A., editors. *Causation and Counterfactuals*. Cambridge, MA: MIT Press, pp. 225–276.

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

Hanson, N. R. (1958). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge, UK: Cambridge University Press.

Hebb, D. O. (2002). *The Organization of Behavior: A Neuropsychological Theory*. Mahwah, NJ: Lawrence Erlbaum.

Hodkinson, I., & Reynolds, M. (2007). Temporal logic. In Blackburn, P., van Benthen, J., and Wolter, F., editors. *Handbook of Modal Logic*. Amsterdam, The Netherlands: Elsevier, pp. 655–720.

Hume, D. (2000). *An Enquiry Concerning Human Understanding: A Critical Edition*, English edition, edited by Beauchamp, T. L. New York: Oxford University Press.

Kant, I. (1933). *Critique of Pure Reason*, English edition, 2nd impression, translated by Smith, N. K. Macmillan.

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

Lewis, D. (1973b). *Counterfactuals*. Oxford, UK: Blackwell.

Lewis, D. (1986). Events. *Philosophical Papers*, **2**, 241–269.

Lewis, D. (2004). Causation as influence. In Collins, J. D., Hall, E. J., and Paul, L. A., editors. *Causation and Counterfactuals*. Cambridge, MA: The MIT Press, pp. 75–106.

Lin, F. (1996). Embracing causality in specifying the indeterminate effects of actions. In *Proceedings of the 13th National Conference on Artificial Intelligence*. CA: Morgan Kaufmann, pp. 1985–1991.

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, **2**(4), 245–264.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. New York: Oxford University Press.

McCain, N., & Turner, H. (1997). Causal theories of action and change. In *Proceedings of the 14th National Conference on Artificial Intelligence*. CA: Morgan Kaufmann, pp. 460–465.

Mellor, D. H. (1995). *The Facts of Causation*. London: Routledge.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Schwind, C. (1999). Causality in action theories. *Electronic Articles in Computer and Information Science*, **4**(4), 27–50.

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press

Sutton, R. S. (2001). *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press.
Thielscher, M. (1997). Ramification and causality. *Artificial Intelligence*, **89**(1–2), 317–364.
Venema, Y. (1993). Derivation rules as anti-axioms in modal logic. *The Journal of Symbolic Logic*, **58**(3), 1003–1034.
von Kutschera, F. (1993). Causation. *Journal of Philosophical Logic*, **22**(6), 563–588.
Williamson, J. (2006). Causality. In *Handbook of Philosophical Logic* (second edition), Vol. 14. The Netherlands: Springer, pp. 95–126.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.
Xu, M. (1997). Causation in branching time (i): Transitions, events and causes. *Synthese*, **112**(2), 137–192.
Zhang, D., & Foo, N. (2001). Epdl: A logic for causal reasoning. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence*. CA: Morgan Kaufmann, pp. 131–136.

INSTITUTE OF LOGIC AND COGNITION
SUN YAT-SEN UNIVERSITY
GUANGZHOU 510275, CHINA
*E-mail*: liuhu2@mail.sysu.edu.cn, wxflogic@gmail.com

https://doi.org/10.1017/S1755020312000299 Published online by Cambridge University Press