Causation and the Semantics of Counterfactuals

MSc Thesis *(Afstudeerscriptie)*

written by

Haitao Cai
(born February 18th, 1989 in Anhui, China)

under the supervision of **Prof Dr Michiel van Lambalgen**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
Sep 21, 2012	Prof Dr Michiel van Lambalgen
	Prof Dr Frank Veltman
	Prof Dr Benedikt Löwe
	Dr Katrin Schulz

Institute for Logic, Language and Computation

Abstract

Semantics of counterfactuals is normally developed according to the principle of similarity, and the key point is to specify the notion of relative similarity. It can be seen in specific examples that causation plays a crucial role in determining the independence of particular facts and thus also in the measure of relative similarity.

Counterfactual account of causation is briefly reviewed. Some counterexamples are provided to illuminate its inherent difficulty, which justifies the proposal that causation underlies the semantics of counterfactuals rather than the converse.

Event calculus is introduced in order to facilitate discussions on causation and the semantics of counterfactuals. Two notions of causation are defined formally under the framework of event calculus by making use of timing.

Given the mechanism of identification of causation, the semantics of counterfactuals is defined in terms of relative similarity. It's argued that the examples about duchess that are used to support the proposal of epistemic reading actually don't work. The example about Kennedy is also analyzed without appealing to epistemic reading.

Contents

1 Introduction

Counterfactual conditionals (for short, counterfactuals) are conditionals whose antecedents are false, or equivalently, contrary to reality. They can be of various syntactic forms, for example

(1) If I were you, I would have a break first.

(2) If Jack had received the mail, he would have replied you.

As long as the speaker isn't speaking to himself, he cannot be the listener, and it implies that the antecedent of sentence (1) is false. To let sentence (2) qualify as a counterfactual, it would have to be the case the Jack hadn't received the mail. So counterfactuals don't describe the state of affairs of the actual world, at least not directly; instead, they are about those possible situations or worlds where their antecedents hold.

Counterfactuals are widely known to be semantically vague since it often happens that people feel hard to reach an agreement on the truth of a counterfactual. The following sentence may well be a good illustration.

(3) If kangaroos had no tail, they would topple over.

It has been figured out by zoologists that kangaroos's tails are crucial for their body balance, therefore lots of people would like to think sentence (3) is true. On the other hand, almost all mammals have got their own ways to keep body balance through millions of years of evolution. Then it's also reasonable to conjecture that kangaroos would have evolved in such a way that they could avoid toppling over even if without tails.

Despite the vagueness surrounding counterfactuals, Lewis believes that it's possible to give a clear account of the truth conditions of them ([7], p.1). Stalnaker [13] and Lewis [7] propose the truth condition of counterfactuals in terms of relative similarity between possible worlds

- A counterfactual conditional is true at an evaluation world w if its consequent holds at all those possible worlds where the antecedent holds and which are most similar (or equivalently, closest) to w.

This proposal, which can be called the *principle of similarity*, is usually adopted as the fundamental principle underlying the semantics of counterfactuals at least by those who are working under the framework of possible world semantics. The main point to be settled in the semantics of counterfactuals based on the principle of similarity, as can be seen, is the *similarity* at issue. Once there is a specific standard of similarity between possible worlds, it would be straightforward to identify those worlds determining the truth of counterfactuals.

The most naive execution of the principle of similarity might be as follows. Denote each counterfactual as $\varphi \hookrightarrow \psi$ where φ is the antecedent and ψ is the consequent.

(i) Each possible world is identified as a function assigning $i \in \{0, 1\}$ to each atomic sentence, or equivalently, a possible world is a maximal consistent set of $\langle p, i\rangle$ where p is an atomic sentence and $i \in \{0, 1\}$.

(ii) Let w be the actual world, relative similarity relation $<_w$ is determined merely by the membership between possible worlds's intersections with w, that is, given two possible worlds w' and w'', we say w' is more similar to w than w'' is (formally, $w' <_w w''$) iff $w'' \cap w \subsetneq w' \cap w$.

(iii) Thus, a closest $[\![\varphi]\!]$-world w' is such that φ holds at w' and there is no world w'' such that $w'' \in [\![\varphi]\!]$ and $w'' <_w w'$[1].

This execution often gives rise to counterintuitive predictions in analysis of specific examples. Consider the following situation:

> A glass was put on the table. Yesterday, it fell to the ground. The ground under the table was very hard and the glass was fragile, so the glass broke.

Then, shall we accept the following counterfactual?

(4) If the glass hadn't fallen to the ground, it would have been broken.

Intuitively, we wouldn't; however, if we pursue the extreme similarity between the actual world and counterfactual worlds and thus follow the naive semantics of counterfactuals, we will just abandon the proposition that the glass fell to the ground and suppose the contrary of it without abandoning the proposition that the glass was broken. The counterfactual world would be highly similar to the actual one except that they differ in the issue if the glass fell to the ground. Despite the extreme similarity on the surface, we will feel uncomfortable if we have to accept (4) according to this selection of 'closest' counterfactual world.

How can we imagine that the glass would have been broken even if it hadn't fallen to the ground? To fill this gap, one might need to put some other event e (which doesn't happen in the actual world; otherwise we would have to modify the original situation described above) into the counterfactual world and make it responsible for the glass's being broken. Thus, apart from the event e, it needs to be supposed that e caused the glass's break. Obviously, it would be even more eccentric if the glass broke without any cause.

It doesn't accord with people's normal reasoning patterns if backup causes can be added into the situation at issue in such a free way; otherwise, counterfactual reasoning would be about questions how to maintain some particular facts if something actual

[1] Stalnaker [13] stipulates that there is a unique closest $[\![\varphi]\!]$-world given the evaluation world and the antecedent φ. However, numerous examples show that there can be more than one closest $[\![\varphi]\!]$-world, which is also justified by the theories proposed by Lewis [7], Veltman [17], etc.

had been different, but what people actual do in counterfactual reasoning is inquiring what would (have to) have happened if something actual had been different.

Another undesirable consequence of the naive semantics is, the actual world and all closest counterfactuals worlds agree on the truth value of the consequent as long as the antecedent and the consequent are logically independent, that is, the truth of the antecedent and that of the consequent are determined by two disjoint sets of atomic sentences. This consequence is in no way acceptable since it would imply that no connection between the antecedent and the consequent except by sharing determining atomic sentences.

It seems that something important is missed in the comparison of the similarity between those antecedent-worlds and the actual one. Actually, the glass broke *because* it fell to the ground. To make the counterfactual assumption '*If the glass hadn't fallen to the ground*', it would be more natural if we abandon the proposition that the glass broke. By analysis of more examples, we are likely to get the following conclusion:

> 'In other words, similarity of particular fact is important, but only for facts that do not depend on other facts. Facts stand and fall together. In making a counterfactual assumption, we are prepared to give up everything that depends on something that we must give up to maintain consistency. But we want to keep in as many independent facts as we can.' ([17], p. 164-165)

Then it is of particular interest and importance to specify the nature of the dependence between facts. It is stated in Schulz [12] that dependence/independence of facts is determined by causation, in other words, independent facts are exactly those which don't have causes in the situation at issue. This point could be supported by analysis of numerous examples, since the truth of a counterfactual can be known when enough information about causation is available.

The example above is a simple but good illustration. The glass's break is caused by its falling to the ground and is thus dependent on the latter fact, so the former is abandoned when it's supposed counterfactually that the glass didn't fall to the ground.

A more complicated example is about neuron networks. Each arrow represents a stimulant signal while a blob at the end of a line represents an inhibitory signal. Shaded circles represent neurons that fire. Every successive neuron fires if and only if it receives a stimulant signal without being inhibited. Initially, neither neuron 2 nor 3 is inhibited.

Neuron 2 fires because it's stimulated by the signal from neuron 1. Neuron 4 inhibits neuron 3 from firing and thus protects neuron 2 from being inhibited, which is also in a sense a cause of neuron 2's firing. Then if neuron 4 hadn't fired, neither would neuron 2 have.

The task of constructing and motivating the semantics of counterfactuals is divided as follows.

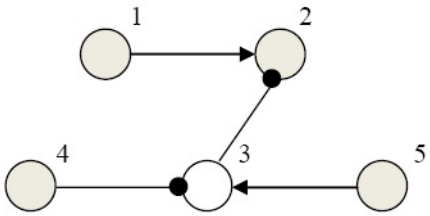


Figure 1

Section 2: The counterfactual account of causation proposed by Lewis and a series of improvements made by advocates of this approach are introduced. But it can be seen in many counterexamples that there is a gap between the differentiating power and the big complication of causal structures. Considering the intimate relations between causation and counterfactuals, this observation justifies the semantics of counterfactuals formulated in terms of causation.

Section 3: A version of event calculus is proposed to facilitate discussions on causation and semantics of counterfactuals. The focus of the event calculus here is on semantic derivation, in which the operation of completion provides information about law-like correlations between atomic facts. Moreover, for the basis of event calculus in physics, some further constraints are imposed to force the finiteness of changes.

Section 4: Following Hall's proposal, two notions of causation are formally defined whose main difference is about the qualification of absence as causal relata. Timing plays a crucial role in identification of causation. It is pointed out that linguistic expressions of causation involves pragmatic factors and is thus different from causal history.

Section 5: The semantics of counterfactuals is given on the basis of causal identification. It is proposed that the notion of causation underlying the semantics is causal contribution. The relative similarity relation is defined in terms of inclusion of overlap with the basis of the model in question. Schulz's theory of epistemic reading of conditionals is briefly reviewed and it's argued that this theory is likely to sacrifice belief's basis in reality. The example about Kennedy that is used to support epistemic reading shows that there is an asymmetry between the weight of normal atomic facts and that of the source X of contingency in comparison of relative similarity.

2 Counterfactual account of causation

Lewis [6] proposes his influential theory of causation in terms of counterfactuals in which counterfactuals are stipulated to be more primitive than causation is. Specifically, Lewis aims at an account of the semantics of those causal claims of the form *'c is a cause of c'*. Under Lewis's stipulation, the truth conditions of counterfactuals are presupposed to be prior to causation. For two events c and e, c is said to depend causally on e iff

both of the following two counterfactuals are true

(5) if c were to occur, so would e;

(6) if c were not to occur, neither would e.

In the opinion of Lewis, causation is not the same as but can be implied by counterfactual dependence among actual events. For actual events c and e, c is a cause of e iff there is a finite chain $c, d_1, \ldots, d_n, e$ of actual events such that each event except for c depends causally on the one immediately before it. Since c is actual, the closest world where c occurs is just the actual world itself. For e is also an actual event, it holds trivially that if c were to occur, so would e. Therefore, the finite chain leading from c to e in Lewis's definition of causation just needs to be such that (6) holds for every pair of adjacent events.

This original proposal is certainly far from being satisfactory even in the eyesight of those theorists who support it. One of the most perspicuous problem is resulted from backtracking counterfactuals. Suppose that e causes c, the worlds most similar to the actual world where c doesn't occur are those in which e doesn't either. More generally, regardless the possibility of disabling conditions, the closest worlds where the effect doesn't occur are those in which the cause doesn't, either. Then it follows by definition that sentence (6) is true, which implies that c is a cause of e. However, it has been assumed that e causes c rather than the converse.

One of most efficient solutions to the reverse between causes and effects is to syntactically preclude backtracking counterfactuals. Specifically, Lewis [8] states that backtracking counterfactuals should be of special syntactic form as follows, if they can be true

- if it had been the case that ψ, it would *have to* have been the case that φ.

With this restriction, (6) wouldn't hold if e is a cause of c since e must temporally precede c. Nevertheless, restricting counterfactuals to the family of non-backtracking ones, counterfactual analysts remain far from completing a counterfactual account of causation because a series of particular causal structures still pose thorny problems for them. One of the most famous difficulties troubling counterfactual analysis is about *preemption*.

It often happens that more than one cause is sufficient to bring about a particular effect though only one of them actually does. Suppose there are two events c_1 and c_2 both of which can lead to event e, but it's c_1 that actually causes e while c_2 doesn't.

Following Lewis's definition of causal dependence, e depends on neither c_1 nor c_2 since either e_1 or c_2 would cause e in case the other were absent. The absence of causal dependence between c's and e inhibits us from identifying c_1 as a cause of e if we follow Lewis's original version of counterfactual account.

Lewis's strategy of coping with this type of difficulties starts with the completion of the causal chain leading from c_1 to e. Without loss of generality, suppose the causal chain is $c_1, d_1, \ldots, d_n, e$, then it is claimed that each event in this chain depends causally on the one immediately before it. The objection is: if d_n had been absent, c_1 would have been absent and thus c_2 would have caused e. His reply to this objection can be mainly divided into two steps.

(a) Even if d_n hadn't occurred, $c_1, d_1, \ldots, d_{n-1}$ would have occurred;

(b) Regardless of the absence of d_n, c_1 would still have interfered with c_2. As a result, c_2 failed to cause e and thus e is also absent.

Step (a) is an alternative way to express the restriction of counterfactuals to the family of non-backtracking ones, namely, in making the counterfactual assumption that an event d hadn't occurred, all facts about events occurring earlier than d are fixed. This stipulation might well be reasonable for those who accept Lewis's claim that backtracking counterfactuals should be of special syntactic forms since only non-backtracking counterfactual reasoning is working here.

However, it's extremely hard to give a justification for step (b) if not impossible. In those counterfactual worlds where c_1 happens and fails to cause d_n in such a way that the causal chain is cut off after d_{n-1}, c_1 would probably have lost the power to interfere with c_2.

An example cited in lots of articles on causation is about two kids who threw stones to a bottle. Suzy threw a stone to the bottle first and then Billy did. Suzy's stone shattered the bottle, so the actual cause of the bottle's shatter is Suzy's throw rather than Billy's. So the reason why Billy's throw didn't cause the bottle's shatter is that Suzy's throw had already shattered the bottle, and that's exactly how Billy's throw was preempted by Suzy's from shattering the bottle. This observation indicates that Billy's throw would no longer be preempted by Suzy's as long as Suzy's throw fails to shatter the bottle.

It's possible to imagine some backup interfering factor which would have made Billy's throw fail to shatter the bottle, but this strategy is rather poor for the same reason as that why we reject the naive semantics of counterfactuals.

There are a series of efforts into saving counterfactuals from the threats posed by preemptions. Among those replies, one that has drawn wide attention is made by focusing on timing of effects, for instance, Paul [10]. Specifically, for two actual events c and e, we say that c is a cause of e iff one of the following holds under the counterfactual assumption that c hadn't occurred

(a) e wouldn't have occurred;

(b) e would have occurred at some time later than the time when it actually occurred.

For the example about bottle shatter, it does work. In the closest worlds where Suzy's throw is absent, the bottle's shatter would happen latter than it actually does though the bottle couldn't avoid being shattered. But the elimination of problems is just an illusion since the causal structure can be designed more elaborately to remove the temporal difference between actual causes and backups. A simple instance is provided by Collins, Hall and Paul [4].

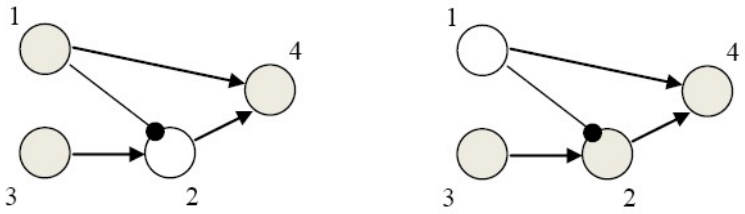


Figure 2

Every time neuron 1 fires, one signal is received by neuron 2 which inhibits its firing even though it's stimulated by the signal from neuron 3, and the other signal from neuron 1 is received by neuron 4 which stimulates neuron 4 to fire. The actual case of the neuron networks is illustrated by the left half of Figure 2. Suppose the neuron network is structured so well that had neuron 1 not fired, the signal from neuron 2 would have been received by neuron 4 at exactly the same time as the signal from neuron 1 actually does. What's more, there is no difference in the temporal length between neuron 4 receiving the signal from either neuron 1 or neuron 2 and its firing.

Then, the actions of neurons in the closest possible world where neuron 1 hadn't fired can be illustrated by the right half of Figure 2. Under the absence of neuron 1's firing, the signal from neuron 3 succeeded in stimulating neuron 2 which caused neuron 4 to fire. For the temporal coincidence between neuron 4's actual firing and that in the counterfactual worlds, the further constraint imposed on the timing of the effect doesn't help us identify the causation between the signal from neuron 1 and the firing of neuron 4.

Another attempt to rescue counterfactual analysis from the threat of preemption is provided by Lewis [9] which can be briefly formulated as follows

- causes can make substantially more notable differences to the way in which effects occur than non-causes can.

Applied to the example about bottle's shattering, the bottle would have been shattered by Suzy's stone in a different way if Suzy had thrown the stone in a somewhat different manner while holding Billy's throw fixed, e.g. Suzy aimed at the neck rather than the body of bottle. On the contrary, even if Billy had thrown the stone in a different manner, there would have been very little difference in the way that the bottle shattered.

Then the contrast between the alterations to bottle's shattering effected respectively by adjustments of Suzy's throw and those of Billy's throw, according to Lewis's proposal, is sufficient for the conclusion that Suzy's throw is a cause of the bottle's shattering while Billy's isn't.

It's obvious that causes often have stronger power to influence the way that effects occur than non-causes do. But it's not always the case. Suppose that the bottle to be shattered is entirely surrounded by a special gravity field such that any stone getting into it would be shot toward the body of bottle with the same momentum, which would eliminate all differences in the manner of bottle's shattering that would have been made by alterations of Suzy's throw if without the existence of the special gravity field. Then Suzy's throw would no longer own advantages over Billy's throw in influencing the way that the bottle is shattered.

It's rather hard to refute counterfactual analysis of causation in such few pages, but it might well have been shown that there is a gap between the somehow poor differentiating power of counterfactuals and the huge complication of possible causal structures. Specifically, the counterfactual approach usually distinguish causes from non-causes by the comparison between their potential to make differences in the presence/absence of effects or the manners that effects occur, but it's probable that some elaborate causal structures can be provided in which the differences concerning effects resulted from causes and non-causes are trickily removed.

On the contrary, given a specific causal structure, the truth of a specific counterfactual concerning the causal structure at issue can usually be evaluated specifically, as is to be seen in Section 5. Therefore, it is likely to be reasonable to regard causation as more primitive than counterfactuals are.

3 Event calculus

To conduct more efficient discussions on causation and counterfactuals, we'd better employ formal tools by which our ideas could be illuminated, examined and justified in a more specific way.

A version of *event calculus* is proposed by van Lambalgen [15] to shed light on the importance of planning to natural language semantics. This version of event calculus, as would be seen, is a very fine-grained framework for characterizations of occurrences of events and changes of properties which are exactly what underlie analysis of causation and counterfactuals. Despite that planning is not the focus of the discussions here, event calculus would greatly facilitate the keen capturing of causation.

3.1 Primitives and axiomatization

Formally, event calculus requires a first-order logic with more abundant primitives while a standard first-order logic has a single domain of individuals.

- ◦ real numbers, by which we could represent time and variable quantities such as distance, energy level, volume, etc;
- ◦ fluent types, possibly with parameters ranging over real numbers, can represent ascriptions of properties to individuals at particular time;
- ◦ event types, when with fixed time, mark the beginning and end of properties.

There are three disjoint domains for the three sorts of primitives above respectively, $\mathbb{R}$ for real numbers, D_e for event types and D_f for fluent types. Real numbers have the normal structure $\langle \mathbb{R}, <; +, \times, 0, 1 \rangle$. It should be noted that the real structure here is characterized by Zermelo-Fraenkel set theory of reals rather than the first-order theory of reals. Properties, expressed by fluents, are primitive rather than interpreted as its extension. The legitimacy of this stipulation can be justified by a series of basic observations, for instance, two distinct properties can share the same extension.

Fluent types can be parameterized, for instance, the volume of water in a pool, which can be noted as $vol(x)$ where x is a variable over real numbers. In the process of filling water into the pool, the volume of water in the pool is changing continuously before the pool is full, so there would need to be predicates marking the initiation as well as the termination of the continuous change. An appropriate argument of the predicate is the abstract fluent $vol()$ for at least two reasons

(i) although $vol(u)$ and $vol(u')$ are two different fluent types as long as $u \neq u'$, it's the case that both of them represent the *volume* of water in the pool;

(ii) it's more succinct to use fluent types with parameters; otherwise, to mark the initiation of a continuous change, we would probably have to use the set of all fluent types without parameters which represent the volume of water in the pool, as an argument of the predicate.

The three sorts of primitives serve as arguments of predicates in event calculus, and most predicates involve not only event types or fluent types but also real numbers which represent time or more specifically, time point at issue. The event calculus aims at giving an account of causation with its basis in physics (van Lambalgen [15], p. 43), so timing is crucial. In each particular causal chain, what is involved are those event tokens and fluent tokens each of which can be given by assigning a specific time point to an event/fluent type.

Among the predicates of event calculus, there needs to be one expressing that an event happens (or equivalently, occurs) and another expressing that an object has a particular property

- $Happens(e,t)$ / $Hap(e,t)$
- $HoldsAt(f,t)$ / $Hol(f,t)$

In natural language, events can have durations, but in event calculus those events with durations are decomposed into (i) instantaneous events, (ii) fluents and (iii) changes of fluents. For example, an event of crossing a street is composed of the event of starting to move, the state of being moving, the continuous change of distance from the start point and the event of reaching the other side of the street, etc.

The two predicates can be related with each other in such a way that an fluent f is initiated or terminated by an event e

- $Initiates(e,f,t)$ / $Ini(e,f,t)$
- $Terminates(e,f,t)$ / $Ter(e,f,t)$

What is expressed by $Ini(e,f,t)$ is f will start to hold immediately after time t if e happens, that is, f is initiated by e at t if $Hap(e,t) \wedge Ini(e,f,t)$. Dually, f is terminated by e if $Hap(e,t) \wedge Ter(e,f,t)$.

Moreover, causation can also be due to a continuous force, for instance, heating and boiling a pot of water, the change of distance as a result of continuous moving, etc.

- $Trajectory(f,t,f',d)$ / $Tra(f,t,f',d)$
- $Releases(e,f,t)$ / $Rel(e,f,t)$
- $Fixes(e,f,t)$ / $Fix(e,f,t)$

$Tra(f,t,f',d)$ expresses that if the fluent f holds between time t and $t+d$ (more strictly, during the open interval $(t,t+d)$), then f' will start to hold at $t+d$. f' often has a parameter over reals, e.g. the volume of water in a pool, the temperature of water, the distance of movement.

Rel expresses the initiation of continuous changes of fluents such that those changes continue without the happening of events. For example, moving can cause continuous changes of distance. Generally, $Rel(e,f,t)$ says that f will start to change continuously if e happens at t. As is argued by van Lambalgen [15], *Rel* is necessary for the consistency between the two types of causes corresponding to Ini and Tra respectively.

> 'Cause as instantaneous change leads to one form of inertia: after the occurrence of the event marking the change, properties will not change value until the occurrence of the next event. This however conflicts with the intended notion of continuous change, where variable quantities may change their values without concomitant occurrences of events.' ([15], p. 44)

Dually, $Hap(e,t) \wedge Fix(e,f,t)$ implies that the continuous change of f will stop at t.

Moreover, a fluent can hold initially for we are usually considering the causal history during a period of finite length.

- $Initially(f)$ / $Ini(f)$

A so-called *principle of inertia* consists in the foundation of event calculus: no change of fluents occur without cause. In other words, no spontaneous change of fluent occur ([15], p. 42). It follows that a fluent holds as long as no interfering event happens after it's initiated.

- $Clipped(t, f, t')$ / $Cli(t, f, t')$
- $Declipped(t, f, t')$ / $Dec(t, f, t')$

$Cli(t, f, t')$ expresses that there is no terminating event or continuous change of f between time t and t'. Dually, $Dec(t, f, t')$ expresses that there is no initiating event or fluent between t and t'.

Fixing D_e and D_f, not all atomic formulas built from the elements of them are of importance; instead, it's often the case that only part of them matter. For instance, we seldom consider if a collision between two balls can terminate the influence of gravity though there are lots of situations where we think of both the collision between two balls and the influence of gravity. Denote the set of atomic formulas at in question by Dom. Note that all atomic formulas in Dom are built from those elements of D_e and D_f and real variables, and no variable over events or fluents occur in any element of Dom.

The axioms of the event calculus given here are developed based on those in [15]. These axioms will hopefully apply to a wider range of examples[2]. All variables over real numbers are supposed to be universally quantified and all event or fluent types are arbitrary elements of D_e and D_f.

AXIOM 1 $Inl(f) \rightarrow Hol(f, 0)$

AXIOM 2 $Hol(f, t) \wedge t < t' \wedge \neg Cli(t, f, t') \rightarrow Hol(f, t')$

AXIOM 3 $Hap(e, t) \wedge Ini(e, f, t) \wedge t < t' \wedge \neg Cli(t, f, t') \rightarrow Hol(f, t')$

AXIOM 4 $\forall s(t < s < t + d \rightarrow Hol(f', s)) \wedge Tra(f', t, f, d) \rightarrow Hol(f, t + d)$

AXIOM 5 $Hap(e, t) \wedge Ter(e, f, t) \wedge t < t' \wedge \neg Dec(t, f, t') \rightarrow \neg Hol(f, t')$

AXIOM 6 $(t \leq s < t' \wedge Hap(e, s) \wedge Ter(e, f, s)) \vee (s < t' \wedge Hap(e, s) \wedge Rel(e, f, s) \wedge t < t' \wedge \neg(\bigvee_{Fix(e', f, r) \in Dom} \exists r(s < r \leq t \wedge Hap(e', r) \wedge Fix(e', f, r)))) \rightarrow Cli(t, f, t')$

AXIOM 7 $(t \leq s < t' \wedge Hap(e, s) \wedge Ini(e, f, s)) \vee (t < s + d \leq t' \wedge \forall r(s < r < s + d \rightarrow Hol(f', r)) \wedge Tra(f', s, f, d)) \rightarrow Dec(t, f, t')$

[2] Axiom 4 in [15] presupposes that any continuous influence, as a cause, must be initiated by an event. However, it's also possible that the continuous influence is itself the effect of another fluent, so the AXIOM 4 above is of a stronger and more general form.

Axiom 1 holds as a consequence of the interpretation of Inl. Axiom 2 formulates the principle of inertia. Axiom 3 says that a fluent f holds at t' if f is initiated by an event e at t and no interfering factor occurs between t and t', that is, $\neg Cli(t, f, t')$, then f will also holds at t'. Dually, we have Axiom 5. Axiom 4 is similar to Axiom 3 except that Axiom 4 characterizes a fluent f initiated by another fluent f'. Axiom 6 defines the predicate Cli: both terminating events and continuous changes of f can interfere with the state that f holds, while Axiom 7 defines Dec.

3.2 The semantic derivation

Similar to other logical theories, it's also a crucial issue in event calculus to specify the mechanism of deriving all the consequences from a family of specific premises. Different from [15], we will be concentrated on semantic derivations only.

Definition 3.1. An *atomic fact* is an atomic sentence gained by replacing all of the variables in a possibly negated atomic formula by constants.

Definition 3.2. A *state* $S(t)$ at time t is a first order formula built from

(1) literals of the form $(\neg)Hol(f, t)$, for t fixed and possibly different f;
(2) formulas in the language of the structure $\langle \mathbb{R}, <, +, \times, 0, 1 \rangle$.

A *causal law* is a sentence of one of the following forms

(1) $S(t) \rightarrow Ini(e, f, t)$ or $Ini(e, f, t)$
(2) $S(t) \rightarrow Ter(e, f, t)$ or $Ter(e, f, t)$
(3) $S(t) \rightarrow Hap(e, t)$
(4) $S(t, d) \rightarrow Tra(f, t, f', d)$ or $Tra(f, t, f', d)$
(5) $S(t) \rightarrow Rel(e, f, t)$ or $Rel(e, f, t)$
(6) $S(t) \rightarrow Fix(e, f, t)$ or $Fix(e, f, t)$

where $S(t)$ (more generally $S(t, d)$) is a state at time t and all event types and fluent types involved are constants rather than variables.

A causal law of the form $S(t) \rightarrow Ini(e, f, t)$ expresses that under circumstances where $S(t)$ is satisfied, the event e can initiate the fluent f at t. Intuitively, the truth value of $Ini(e, f, t)$ is uniquely determined by the relevant properties at time t, so in application $S(t)$ is assumed to be about the truth of those components $(\neg)Hol(f, t)$ with the fixed t. Similarly for Ter, Hap, Rel and Fix. As for Tra, it expresses the initiating effect of a continuous influence which holds during $(t, t + d)$, so $S(t, d)$ should be about the truth values of those fluents involved in $S(t, d)$ during $(t, t + d)$.

Atomic facts of the forms $(\neg)Ini(e, f, t)$, $(\neg)Ter(e, f, t)$, $(\neg)Tra(f, t, f', d)$, $(\neg)Rel(e, f, t)$ and $(\neg)Fix(e, f, t)$ express law-like facts, e.g. $Ini(e, f, t)$ expresses that if e happens at

t, f will start to hold immediately after t (but not at t). It's possible that these facts are unconditional in the situation at issue which is formally represented as causal laws with empty antecedents. This doesn't always mean the causal law does have the universal validity, instead, its unconditional form is usually attained in idealized reasoning process, for example, when an object is in uniform linear motion, its distance from the start point is proportional to the length of time.

Definition 3.3. A *root* $\mathcal{R}$ is a pair $\langle F_P, L_C\rangle$ where F_P is a set of atomic facts of the form $(\neg)Inl(f)$, $(\neg)Hap(e,t)$ or $(\neg)Hol(f,t)$ and L_C is a set of causal laws.

F_P consists of particular facts which are of only three possible forms. Atomic facts of the form $(\neg)Ini(e,f,t)$, $(\neg)Ter(e,f,t)$, $(\neg)Tra(f,t,f',d)$, $(\neg)Rel(e,f,t)$ or $(\neg)Fix(e,f,t)$ don't qualify as particular facts since they express law-like facts. Moreover, atomic facts of the form $Cli(t,f,t')$ or $Dec(t,f,t')$ are absolutely determined by other atomic facts between t and t', and specifically, they are just about the existence of interfering/initiating factors of f.

Definition 3.4. Given the domains D_e, D_f and Dom and a root $\mathcal{R} = \langle F_P, L_C\rangle$, the *completion* $Comp(\mathcal{R})$ of $\mathcal{R}$ is the pair $\langle F_P, L_D\rangle$ where L_D is the set of *definitions* determined by the domains and L_C[3].

(i) For each atomic formula p of the form $Ini(e,f,t)$, $Ter(e,f,t)$, $Tra(f,t,f',d)$, $Rel(e,f,t)$ or $Fix(e,f,t)$,

  (a) if there is $\sigma \to p \in L_C$, then pick all such σ_i, the definition Def(p) of p is $\bigvee_i \sigma_i \leftrightarrow p$;
  (b) if $p \in L_C$, then Def(p) $= p$[4].

(ii) For $Hap(e,t)$,

  (a) if there is $\sigma \to Hap(e,t) \in L_C$, then pick all such σ_i, the definition Def($Hap(e,t)$) of $Hap(e,t)$ is $\bigvee_i \sigma_i \leftrightarrow Hap(e,t)$;
  (b) if there is no $\sigma \to Hap(e,t) \in L_C$, then Def($Hap(e,t)$) is undefined.

[3]Every variable in a definition is supposed to be universally quantified, but in the following formalizations, universal quantifies are omitted.

[4]$\sigma \to p$ and p can't both belong to L_C since a causal law shouldn't be both conditional and unconditional in the same situation

(iii) Fix $f \in D_f$, let

$$
\begin{aligned}
TerCli_e(t,t') &= \exists s(t \le s < t' \wedge Hap(e,s) \wedge Ter(e,f,s)) \\
RelCli_e(t,t') &= \exists s(s < t' \wedge Hap(e,s) \wedge Rel(e,f,s)) \wedge t < t' \wedge \\
&\quad \neg(\bigvee_{Fix(e',f,r) \in Dom} \exists r(s < r \le t \wedge Hap(e',r) \wedge Fix(e',f,r))) \\
IniDec_e(t,t') &= \exists s(t \le s < t' \wedge Hap(e,s) \wedge Ini(e,f,s)) \\
TraDec_{f'}(t,t') &= \exists r \exists d(\forall s(r < s < r+d \rightarrow Hol(f',s)) \wedge Tra(f',r,f,d) \wedge t < r+d \le t') \\
InlHol_f(t) &= Inl(f) \wedge \neg Cli(0,f,t) \\
IniHol_e(t) &= \exists t'(Hap(e,t') \wedge Ini(e,f,t') \wedge t' < t \wedge \neg Cli(t',f,t)) \\
TraHol_{f'}(t) &= \exists t' \exists d(\forall s(t' < s < t'+d \rightarrow Hol(f',s)) \wedge Tra(f',t',f,d) \wedge t \ge t'+d \wedge \\
&\quad \neg Cli(t'+d,f,t))
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\mathrm{Def}(Cli(t,f,t')) &= (\bigvee_{Ter(e,f,t) \in Dom} TerCli_e(t,t')) \vee (\bigvee_{Rel(e,f,t) \in Dom} RelCli_e(t,t')) \\
&\quad \leftrightarrow Cli(t,f,t') \\
\mathrm{Def}(Dec(t,f,t')) &= (\bigvee_{Ini(e,f,t) \in Dom} IniDec_e(t,t')) \vee (\bigvee_{Tra(f',t',f,d)} TraDec_{f'}) \\
&\quad \leftrightarrow Dec(t,f,t') \\
\mathrm{Def}(Hol(f,t)) &= InlHol_f(t) \vee (\bigvee_{Ini(e,f,t) \in Dom} IniHol_e(t)) \vee (\bigvee_{Tra(f',t',f,d) \in Dom} TraHol_{f'}(t)) \\
&\quad \leftrightarrow Hol(f,t)
\end{aligned}
$$

The definition of each form of atomic formula is intuitive. For instance, $Cli(t,f,t')$ holds if (i) there is an event e happening and terminating f between t and t' or (ii) there is temporal overlap between (t,t') and a continuous change of f. Similarly for $Dec(t,f,t')$. $Hol(f,t)$ holds if (i) f holds initially; (ii) f is initiated by some event or (iii) f is initiated by another fluent.

The crucial step in constructing the definition of an atomic formula is *bi-conditionalization*. The intuition underlying such an operation is that causal reasoning is usually conducted in a somehow closed world. Particularly, it can be summarized as follows

(a) the truth of each atomic formula is determined by a limited set of fluents or events;

(b) if all causes (if there are) fail to force a particular effect, this effect will not occur.

To satisfy (a), it would have to be the case that both D_e and D_f are of finite cardinalities if each maximal family of fluent types which differ only in their parameters is counted as one.

(b) expresses a version of *negation as failure* ([16], p. 33). If without a world's being closed, causal reasoning would be extremely hard to conduct since the subject would always need to consider what would happen if an event or a fluent is caused by an unknown force. But extension of domains and causal laws are permissible when the subject acquires or retrieves more factors relevant to the situation in question.

Analogously, it ought to be stipulated that for all atomic formulas $p \in Dom$ of the form $Ini(e, f, t)$, $Ter(e, f, t)$, $Tra(f, t, f', d)$, $Rel(e, f, t)$ or $Fix(e, f, t)$, there is a causal law in L_C with the consequent p; otherwise, it would be assumed that p doesn't hold at any time in question since there is no evidence forcing its truth and thus deserves no attention.

Completion provides law-like correlations between each atomic formulas and those directly related to it, so it doesn't apply to elements of F_P since they are *particular facts* which hold specifically without application of laws. The division between particular facts and laws is crucial to the semantics of counterfactuals, which is to be introduced briefly in Section 5.

3.3 Finiteness of changes

If the causal reasoning at issue does have its basis in physics, then it should be assumed that each event happens only finitely many times during a period of finite length. To force the finiteness of occurrences of events, the following two constraints are imposed which say that for each open interval (t, t') of finite length, if an event e happens during (t, t'), then there is an earliest time and a latest time when e happens.

(1a) $\exists s(t < s < t' \wedge Hap(e, s)) \rightarrow \exists t''(t < t'' < t' \wedge Hap(e, t'') \wedge \neg\exists s'(t < s' < t'' \wedge Hap(e, s')))$

(1b) $\exists s(t < s < t' \wedge Hap(e, s)) \rightarrow \exists t''(t < t'' < t' \wedge Hap(e, t'') \wedge \neg\exists s'(t'' < s' < t' \wedge Hap(e, s')))$

Lemma 3.1. Given arbitrary $t_1, t_2 \in \mathbb{R}$ and $e \in D_e$, for all models $\mathcal{M}$ of Constraint (1a) and (1b) it holds that $|\{t \in (t_1, t_2) \mid \mathcal{M} \models Hap(e, t)\}|$ is finite.

Proof. Suppose that there are infinitely many $t \in (t_1, t_2)$ such that $\mathcal{M} \models Hap(e, t)$. We construct a sequence $\delta_{i\in\mathbb{N}}$ as follows. By Constraint (1a), there is a minimal $t \in (t_1, t_2)$ such that $\mathcal{M} \models Hap(e, t)$, let $\delta_0 = t$. Applying Constraint (1a) again and again, let δ_{i+1} be the minimal $t \in (\delta_i, t_2)$ such that $\mathcal{M} \models Hap(e, t)$. For all $i \in \mathbb{N}$, let $S_i = (t_1, \delta_i)$.

Claim 1: there is $u \in (t_1, t_2)$ such that $u \notin \bigcup_{i\in\mathbb{N}} S_i$.

By Constraint (1b), there is a maximal $t_3 \in (t_1, t_2)$ with $\mathcal{M} \models Hap(e, t_3)$. Pick an arbitrary $u \in (t_3, t_2)$, it holds that there is no $t \in (u, t_2)$ such that $\mathcal{M} \models Hap(e, t)$.

On the other hand, by definition of S_i, for all $s \in \bigcup_{i\in\mathbb{N}} S_i$, there is $n \in \mathbb{N}$ such that $s \in (t_1, \delta_n)$. It follows that $u \notin \bigcup_{i\in\mathbb{N}} S_i$.

Let v be the greatest lower bound of $X = \{t \in (t_1, t_2) \mid t \notin \bigcup_{i\in\mathbb{N}} S_i\}$, then for all $t \in (t_1, v)$ we have $t \in \bigcup_{i\in\mathbb{N}} S_i$.

Claim 2: $v \notin \bigcup_{i\in\mathbb{N}} S_i$.

Otherwise, suppose that $v \in \bigcup_{i\in\mathbb{N}} S_i$, then there is $n \in \mathbb{N}$ such that $v \in S_n$, which implies that $v \in (t_1, \delta_n)$. So, for all $t \in (v, \delta_n)$ we have $t \in S_n$ and thus $t \in \bigcup_{i\in\mathbb{N}} S_i$. Pick an arbitrary $v' \in (v, \delta_n)$, it isn't hard to see that v' is also a lower bound of X, contradicting the assumption that v is the greatest lower bound.

Since for all $t \in (t_1, v)$ we have $t \in \bigcup_{i\in\mathbb{N}} S_i$, there is no maximal $t \in (t_1, v)$ with $\mathcal{M} \models Hap(e, t)$, which contradicts (1b). Otherwise, suppose that s is the maximal element. Since for all $t \in (t_1, v)$ we have $t \in \bigcup_{i\in\mathbb{N}} S_i$, there is $n \in \mathbb{N}$ such that $s \in S_n$, so it holds that $v \in (s, \delta_n] \subseteq (t_1, \delta_{n+1}) = S_{n+1} \subseteq \bigcup_{i\in\mathbb{N}} S_i$ because of the maximality of s and the infinity of the sequence $\delta_{i\in\mathbb{N}}$, contradicting Claim 2.

Thus, the supposition made at the beginning of the proof can't hold.

□

Similarly, it should also hold that the time when a fluent holds during a period of finitely length can be divided into finitely many intervals, which could be forced by the following two constraints.

Let $\chi(f, t) = \exists f' \exists s \exists d(\forall t'(s < t' < s + d \to Hol(f, t')) \wedge Tra(f', s, f, d) \wedge t = s + d)$

(2a) $\exists t(t_1 < t < t_2 \wedge Hol(f, t)) \to \exists s'(t_1 < s' < t_2 \wedge \chi(f, s') \wedge \neg\exists r(t_1 < r < s' \wedge \chi(f, r)))$

(2b) $\exists t(t_1 < t < t_2 \wedge Hol(f, t)) \to \exists s'(t_1 < s' < t_2 \wedge \chi(f, s') \wedge \neg\exists r(s' < r < t_2 \wedge \chi(f, r)))$

Since it is stipulated that D_e is of finite cardinality and it has been proved above that each event (type) occurs finitely many times during each interval of finite length, it holds that each fluent (type) is initiated and terminated by events only finitely many times during each finitely long interval. (2a) and (2b) guarantees that each fluent is initiated by continuous forces for only finitely many times during each finitely long interval. Thus we have the following lemma.

Lemma 3.2. Given the finite domains D_e, D_f and Dom and a root $\mathcal{R} = \langle F_P, L_C\rangle$, for arbitrary $(t_1, t_2) \subseteq \mathbb{R}$ and all models $\mathcal{M}$ of $Comp(\mathcal{R})$, (1a), (1b), (2a) and (2b) and all $f \in D_f$, it holds that $\{t \in (t_1, t_2) \mid \mathcal{M} \models Hol(f, t)\}$ equals a finite union of intervals of the form $(r, s]$ or $[r, s]$.

Proof. Omitted. □

Then, given the domains and a root $\mathcal{R} = \langle F_P, L_C \rangle$, what is of interest is the family of atomic facts entailed by $F_P \cup L_D \cup \{(1a), (1b), (2a), (2b)\}$, that is, those atomic facts p such that

$$F_P \cup L_D \cup \{(1a), (1b), (2a), (2b)\} \models p$$

where the entailment $\models$ is only with respect to those models with the normal structure of reals.

3.4 An example

Fill a pool of volume v by a tap with speed 1, the tap was turned off automatically when the bottle is full. It is assumed that the tap is turned on only once (at time 0) during $[0, u]$ where $u > v$. The volume of water in the pool at each instant during $[0, u]$ can be specified.

$D_e = \{on, off\}$, $D_f = \{flow, vol(x)\}$,

$Dom = \{Hap(on, t),\ Hap(off, t),\ Hol(flow, t),\ Hol(vol(x), t),\ Ini(on, flow, t),\ Ter(off, flow, t), Rel(on, vol(), t), Fix(off, vol(), t), Tra(flow, t, vol(x), d), Cli(t, flow, t'),\ Cli(t, vol(x), t)\}$.

Considering that not all atomic sentences gained by replacing the variables in elements of Dom are of importance, there should be a domain AT of atomic sentences. Here, AT contains those atomic facts given by fixing the variables in the elements of Dom in such a way that $\{t, t'\} \subseteq [0, u]$, $\{d, t + d\} \subseteq (0, u]$, $t \leq t'$, $x \in [0, v]$, which will be evaluated. Some atomic facts are of no importance even if they are given by fixing the real variables in the elements of Dom, for instance, we don't consider the question when it holds that the water in the pool is of volume larger than v since the volume the whole pool is only v.

The root $\mathcal{R} = \langle F_P, L_C \rangle$ where

$$F_P = \{Inl(vol(0)), \neg Inl(flow), Hap(on, 0)\} \cup \{\neg Hap(on, t) \mid t \in (0, u]\}$$

and L_C consists of the following causal laws:

- $\neg Hol(flow, t) \rightarrow Ini(on, flow, t)$
- $Hol(flow, t) \rightarrow Ter(off, flow, t)$
- $\neg Hol(flow, t) \wedge Hol(vol(y), t) \wedge y < v \rightarrow Rel(on, vol(), t)$[5]
- $Hol(flow, t) \rightarrow Fix(off, vol(), t)$
- $Hol(vol(y), t) \wedge y + d \leq v \rightarrow Tra(flow, t, vol(y + d), d)$
- $Hol(vol(v), t) \wedge Hol(flow, t) \rightarrow Hap(off, t)$

[5]This causal law indicates that the real variables involving in a state shouldn't be limited to those representing time; instead, real variables can also be parameters of fluent types which are also universally quantified in the completion of a root.

The definitions of the six atomic formulas are then straightforward. Moreover, we have

- $\mathrm{Def}(Cli(t, flow, t')) = \exists s(t \leq s < t' \wedge Hap(off, s) \wedge Ter(off, flow, s)) \leftrightarrow Cli(t, flow, t')$
- $\mathrm{Def}(Cli(t, vol(y), t')) = \exists s(s < t' \wedge Rel(on, vol(), s) \wedge Hap(on, s) \wedge t < t' \wedge \neg\exists r(s < r \leq t \wedge Hap(off, r) \wedge Fix(off, vol(), r))) \leftrightarrow Cli(t, vol(y), t')$
- $\mathrm{Def}(Hol(flow, t)) = \exists t'(Hap(on, flow, t') \wedge t' < t \wedge \neg Cli(t', f, t)) \vee (Inl(flow) \wedge \neg Cli(0, flow, t)) \leftrightarrow Hol(flow, t)$
- $\mathrm{Def}(Hol(vol(y), t)) = \exists t' \exists d(\forall s(t' < s < t' + d \rightarrow Hol(flow, s)) \wedge Tra(flow, t', vol(y), d) \wedge t \geq t' + d \wedge \neg Cli(t' + d, vol(y), t)) \vee (Inl(vol(y)) \wedge \neg Cli(0, vol(y), t)) \leftrightarrow Hol(vol(y), t)$

The following constraint is also added for this story

$$Hol(vol(x), t) \wedge Hol(vol(y), t) \rightarrow x = y$$

which is named as *integrity constraint* in [15] and contributes to a complete structure of events and fluents. Denote this integrity constraint by Ic. Ic says that the volume of the water in the pool is of a unique value at each instant. Let $\mathcal{P} = F_P \cup L_D \cup \{(1a), (1b), (2a), (2b), Ic\}$, we can fix a unique model satisfying $\mathcal{P}$.

By the definition of $\mathrm{Def}(Cli)$, it holds trivially that

$$\mathcal{P} \models \{\neg Cli(t, flow, t) \mid t \in [0, u]\} \cup \{\neg Cli(t, vol(x), t) \mid t \in [0, u] \wedge x \in [0, v]\}$$

By $\mathrm{Def}(Hol(flow, t))$, we have $\mathcal{P} \models \neg Hol(flow, 0)$.

Step by step:
$\mathcal{P} \models Hol(vol(0), 0)$, $Ini(on, flow, 0)$, $\neg Ter(off, flow, 0)$, $\neg Fix(off, vol(), 0)$,

$\mathcal{P} \models \{Tra(flow, 0, vol(d), d) \mid d \in (0, v]\} \cup \{Rel(on, vol(), 0)\}$.

By constraint (1a), if e happens during $(0, v)$, then there must be a minimal $t \in (0, v)$ such that we have $Hap(e, t)$. On the other hand, if no $t \in (t_1, t_2)$ could be such a minimal element, then it must be the case that e never happens during $(0, v)$.

For an arbitrary model $\mathcal{M}$ of $\mathcal{P}$, suppose the minimal element is $v' \in (0, v)$, that is,

$$\mathcal{M} \models \{Hap(off, v')\} \cup \{\neg Hap(off, t) \mid t \in (0, v')\}$$

then it follows that

$$\mathcal{M} \models \{\neg Cli(0, flow, s) \mid s \in (0, v']\} \cup \{Hol(vol(v), v')\}$$

and then

$$\mathcal{M} \models \{Hol(flow, s) \mid s \in (0, v']\}$$

by $\mathrm{Def}(Hol(vol(x), t))$ we have

$$\mathcal{M} \models Hol(vol(v'), v')$$

contradicting the integrity constraint $Hol(vol(x),t) \wedge Hol(vol(y),t) \rightarrow x = y$, which implies there is no such minimal element and thus $\mathcal{M} \models \{\neg Hap(off,t) \mid t \in (0,v)\}$.

Since $\mathcal{M}$ is arbitrary, it holds that

$$\mathcal{P} \models \{\neg Hap(off,t) \mid t \in (0,v)\}$$

which implies that

$$\mathcal{P} \models \{\neg Cli(t,flow,t') \mid (t,t') \subseteq (0,v)\} \cup \{Cli(t,vol(y),t') \mid t < v \wedge t < t' \wedge y \in [0,v]\}$$

and so

$$\mathcal{P} \models \{Hol(flow,t) \mid t \in (0,v]\}$$

$$\begin{aligned}\mathcal{P} \models & \{Hol(vol(t),t) \mid t \in (0,v]\} \cup \{Ter(off,flow,t) \mid t \in (0,v]\} \cup \\ & \{\neg Ini(on,flow,t) \mid t \in (0,v]\} \cup \{Fix(off,vol(),t) \mid t \in (0,v]\} \\ & \{\neg Rel(on,vol(),t) \mid t \in (0,v]\}\end{aligned}$$

Then we have

$$\mathcal{P} \models \{Hap(off,v)\} \cup \{Tra(flow,t,vol(x),d) \mid t + d = x \in (0,v]\}$$

and it follows that

$$\mathcal{P} \models \{Cli(t,flow,t') \mid v \in [t,t') \subseteq [0,u]\} \cup \{\neg Cli(t,vol(y),t') \mid v \leq t \leq t' \wedge y \in [0,v]\}$$

by Def($Hol(vol(x),t)$ and Def($Hol(flow,t)$), we can attain the result we are usually most interested in

$$\mathcal{P} \models \{Hol(vol(v),t) \mid t \in [v,u]\} \cup \{\neg Hol(flow,t) \mid t \in (v,u]\}$$

To complete the model, we have further that

$$\begin{aligned}\mathcal{P} \models & \{\neg Ter(off,flow,t) \mid t \in (v,u]\} \cup \{Ini(on,flow,t) \mid t \in (v,u]\} \cup \\ & \{\neg Rel(on,vol(),t) \mid t \in (v,u]\} \cup \{\neg Fix(off,vol(),t) \mid t \in (v,u]\} \cup \\ & \{\neg Hap(off,t) \mid t \in (v,u]\}\end{aligned}$$

and then

$$\mathcal{P} \models \{\neg Cli(t,flow,t') \mid t \in (v,u]\}$$

By Ic,

$$\mathcal{P} \models \{\neg Hol(vol(x),t) \mid x \neq t \in [0,v] \vee (x \neq v \wedge t \in (v,u])\}$$

and by Def($Tra(flow, t, vol(x), d)$)

$$\mathcal{P} \models \{\neg Tra(flow, t, vol(x), d) \mid x \neq t + d\} \cup \{\neg Hap(off, 0)\}$$

When we want to specify the truth value of an atomic fact p with temporal index other than 0, there are always infinitely many intermediate atomic facts lying between p and those atomic facts with index 0 because of the particular structure of reals. No step-by-step operation has been found which can derive all facts from the initial states and events of the situation in question. This issue is worth intense attention in future studies. Currently, the semantic derivation has to appeal to (1a), (1b), (2a) and (2b), as is illustrated by the example above.

4 Identification of actual causes

Given a structure of events and fluents, it is often a very controversial issue which events and fluents are the causes of another event or fluent. Causation is an abstract relation between events and fluents and covers numerous types of influence, for example, transformations of momentum and energy caused by collisions between objects, heating raises the temperature of a pot of water, etc.

To see part of the complication of the problem about identification of actual causes, we could have a look at Figure 1 again. The firing of neuron 4 doesn't enter the causal chain of the neuron 2's firing, since the former just inhibits the disabling factor of the latter. On the other hand, no one could deny that neuron 4's firing is necessary for the the firing of neuron 2. When one is asked, why does neuron 2 succeed in firing even in the situation where neuron 5 fires, he would probably answer, because neuron 1 fires which stimulates neuron 2, and neuron 3 fires which protects neuron 2 from being inhibited, in other words, neuron 4 contributes to the absence of disabling condition of neuron 2's firing. Then, can neuron 4's firing be counted as a cause of neuron 2's firing?

Speaking more generally, the difficulty of identification of actual causes is often due to the disputation over the qualification of absence as causes and effects.

On one hand, absence is the non-existence of some event or fluent. As Lewis explained

> 'Absences are not events. they are not *anything*: Where an absence is, there is nothing relevant there at all. Absences are bogus entities.' ([9], p.100)

which implies that absences are nothing and it doesn't make sense to distinguish an absence from another. Otherwise, to differentiate two absences, or more generally, two arbitrary entities, there should be some property which can be ascribed to one of them but not the other. However, absence is non-existence in which no property can be found. For instance, it's weird to talk about the difference between absence of juice and that

of coffee. Causation is thought to hold between distinct events or fluents ([9], p. 78), so there cannot be causation involving absences for they are nothing, and admitting causation involving absences will inevitably lead to causation between two absences.

On the other hand, we can say that the light remains on *because* it hasn't been turned off and more generally, a fluent holds because it hasn't been terminated. Also, absence of drinking water causes thirst.

To solve this sort of puzzles, Hall [3] comes up with the idea that there are two concepts of causation. According to the first notion of causation, neuron 2 is *produced* and neuron 3, 4 and 5 are not involved in the production. According to the second notion of causation, all factors relevant to neuron 2's firing are taken into account, among which only neuron 1 and 3 contribute positively to the firing of neuron 2, though in quite different manners.

Even if out of the actual causal chain of the expected effect, those contributing to the prevention of interfering factors are also important. For instance, when designing an experiment, chemists not only think about the chain of reactions producing the target substance but also try their best to employ some other reactions to preclude all those interfering substances. The target substance is attained *because* chemists also conducted the latter ones.

4.1 Causal production

Following the terminology proposed by Hall [3], the first notion of causation is called *production*, and more specifically, causal production or productive causation. It's not easy to define the notion of causal production precisely. In specific situations, the identification of causes of a particular effect can also be a very hard work and inevitably limited by human cognition and technologies.

An attribute of productive causation is, each pair of cause and effect are connected through a chain of exchange of substance and transformation of energy. Transformation of energy doesn't necessarily involve (visible) contact, for example, it can happen through magnetic fields or gravity fields. In a wider sense, people's will, desire and determination can also be involved in causal chains, which is crucial to ethics. Another attribute of productive causation is temporal precedence, that is, causes must precede their effects.

The identification of causation under event calculus is conducted as a version of reductionist approach. Typically, a reductionist approach attempts to reduce facts about causation to two sorts of facts, (1) facts about what happens, or briefly, *categorical facts* and (2) facts about the laws that govern the happening of categorical facts, i.e. *nomological facts* ([4], p. 12). Then, the core task in causal identification is to specify which causal laws, among all causal laws in question, capture the mode or manner in which changes actually happen, since there are often more than one causal law according to which a particular change can occur. For example, both Suzy's throw and Billy's

are sufficient for shattering the bottle while only one of them is the actual cause of the bottle's shattering.

The main objections to reductionist approaches usually aim at showing that the facts taken into account by reductionists are often insufficient to fix the causal facts at issue. A common version of those objections could be as follows.

> The only causal law concerned is: whenever event e happens, it causes one and only one among e' and e''. Moreover, some cause, which we have almost no idea about, can also give rise to e' and it's a similar case for e''. Suppose at some particular moment, e happens and immediately later both e' and e'' are observed to happen. By the causal law, either e' or e'' is caused by e, but it can't be both, so one of them is caused by unknown cause. Given these nomological facts and categorical facts, it's still unsure if e caused e' or e''.

This type of attack can hardly harm the value and plausibility of reductionist approaches since these objections implicitly impose excessive requirements on the power of each particular reductionism, i.e. causal facts should be fixed even when we have only a very limited class of causal laws and categorical facts.

But actually, the hesitation between e' and e'' can naturally be ascribed to the incompleteness of causal laws, that is, we need to further specify the mechanism of the occurrences of e' and e'' given e, through which we will be able to capture the factors determining if e' or e'' follows from e. Then we would have at least two causal laws relevant to the judgment here rather than only one, for instance,

(i) under the condition f holds, e causes e' while
(ii) under the condition that f doesn't hold, e causes e''.

Event calculus is born to be an ideal framework for the formal analysis of causal production for a series of reasons, e.g. temporal continuity is of fundamental importance in event calculus which facilitates and is often necessary for the specification of causation, and specifically, the temporal location of an effect is immediately after its direct causes without any gap.

Definition 4.1. Given domains D_e of events, D_f of fluents, Dom of atomic formulas and AT of atomic facts, a root $\mathcal{R} = \langle F_P, L_C\rangle$ with the completion $Comp(\mathcal{R}) = \langle F_P, L_D\rangle$. Let $\mathcal{P} = F_P \cup L_D \cup \{(1a), (1b), (2a), (2b)\}$, $M = \{(\neg)p \mid p \in AT \wedge \mathcal{P} \models (\neg)p\}$, $FE = \{p \in M \mid p$ is of the form $Hap(e,t)$ or $Hol(f,t)\}$, the *causal production* $\dashrightarrow$ is the intersection of

(a) $FE \times FE$ and
(b) the transitive closure $TC(\succ)$ of $\succ$.

where the binary relation $\succ \subsetneq M \times M$ is defined as follows. For $p, p' \in M$, we say p is a *productive cause* of p' or p participates in the *causal production* of p' if $p \dashrightarrow p'$.

Although $\dashrightarrow$ holds only between atomic facts of the form $Hap(e,t)$ or $Hol(f,t)$, $\succ$ also involves atomic facts of other forms since they are auxiliary and necessary.

(1) For $p(t,d) \in \{Hap(e,t), Ini(e,f,t), Ter(e,f,t), Tra(f,t,f',d), Rel(e,f,t), Fix(e,f,t)\}$ with definition

$$\mathrm{Def}(p(t,d)) = \bigvee_i \sigma_i \leftrightarrow p(t,d)$$

let $p(t_0,d_0)$ be the atomic fact given by substituting constants t_0 and d_0 for the variables t and d in $p(t,d)$.

Then it holds that

$$Hol(f,t_1) \succ p(t_0,d_0)$$

if there is i such that $\mathcal{P} \models \sigma_i(t_0,d_0)$ and $\sigma_i(t_0,d_0) \models Hol(f,t_1)$.

(2) For $Hol(f,t)$ with

$$\mathrm{Def}(Hol(f,t)) = InlHol_f(t) \vee (\bigvee_{Ini(e,f,t') \in Dom} IniHol_e(t)) \vee (\bigvee_{Tra(f',t',f,d) \in Dom} TraHol_{f'}(t)) \leftrightarrow Hol(f,t)$$

such that $Hol(f,t_0) \in M$,

(2.1) if $\mathcal{P} \models IniHol_e(t_0)$ for some $Ini(e,f,t') \in Dom$, then pick the t_1 witnessing the existential quantifier in $IniHol_e$, we have

$$Hap(e,t_1) \succ Hol(f,t_0)$$
$$Ini(e,f,t_1) \succ Hol(f,t_0)$$

(2.2) if $\mathcal{P} \models TraHol_{f'}(t_0)$ for some $Tra(f',t',f,d) \in Dom$, then pick the t_1, d_0 witnessing the existential quantifiers in $TraHol_{f'}$, we have

$$Hol(f',t_2) \succ Hol(f,t_0)$$

for all t_2 such that $t_1 < t_2 < t_1 + d_0$, and

$$Tra(f',t_1,f,d_0) \succ Hol(f,t_0)$$

.

Clause (2) can be interpreted as follows. If a fluent f is initiated by an event e or another fluent f' at time t' which is earlier than t, and no interfering factor occurs between t' and t, then f is caused by e/f'. The absence of interfering factor precludes

any other initiation of f between t' and t which guarantees that e/f' must be an actual cause of f under the principle of inertia. No initiation of f between t and t' can happen because of f holds all the time during that period.

$\succ$ actually covers several types of relations, for example, in clause (2.1) of Definition 4.1, $Hap(e, t_1) \succ Hol(f, t_0)$ represents that e serves as an initiating event of f while $Ini(e, f, t_1) \succ Hol(f, t_0)$ connects f's initiation to those enabling conditions.

$\succ$ also holds between atomic facts of the same temporal index, so it is a relation between different aspects of fluents. For each $p \in \{Ini(e, f, t), Ter(e, f, t), Tra(f, t, f', d), Rel(e, f, t), Fix(e, f, t)\}$, if there is a causal law with the consequent p, then the truth of p at each instant is determined by those fluents's (which constitute the antecedent of the causal laws with consequent p) truth at the same instant or during the period $(t, t+d)$. A plausible interpretation of those p's is, they represent the *functional* aspects of the fluents involved in the antecedents of causal laws.

For example, a causal law of the form $S(t) \rightarrow Ini(e, f, t)$ expresses that at any time t when state $S(t)$ holds, event e can initiate the fluent f; or briefly, the state $S(t)$ enables e to initiate f, as long as e happens at time t. One often characterizes a situation by specifying which fluents hold in the situation and which don't, e.g. there is a high concentration of carbon monoxide in the room; but he can also do it by describing the changes that can occur, e.g. the room is very dangerous since a spark can trigger a fierce explosion in it.

Since the purpose here is identifying those actual causes of particular effects, we don't need to distinguish the different roles formally that the actual causes play. What is important is that causal laws and particular facts tell us that they *do* participate in the productions of those particular effects according to axioms and causal laws.

For all $p \in \{Hol(f, t), Hap(e, t), Ini(e, f, t), Ter(e, f, t), Tra(f, t, f', d), Rel(e, f, t), Fix(e, f, t)\}$, $\neg p$ is never an argument of $\succ$. This stipulation will be natural if Axiom 3, 4 and 5 are thought of as exhausting all patterns of causal production, since $\neg p$ just expresses the absence of a property, an event or potential for changes. *Production* is a vivid name for this notion of causation. Absence can in no means participate in causal production since productions, if interpreted literally, are transformations from (existent) entities to (existent) entities.

To apply the definition of causal production to the account of neuron networks, the mechanism of a neuron's stimulation and firing should be specified. A relatively simple but very probable mechanism is as follows,

(i) the neuron of energy 0 (unexcited state) receives the stimulus from preceding neurons, which initiates the continuous raising of energy level of the neuron at speed 1;

(ii) the continuous raising of energy level leads to the excited state (energy d) of the neuron after a period of length d;

(iii) the neuron, if uninhibited, fires immediately when it goes into the excited state.

Signals from some neurons can be inhibitory for others. So a neuron may receive two kinds of signals, namely, *stimulus* and *inhibition*. Moreover, there is also a process between a neuron's firing and its target neuron's receiving the signal. So there are two fluents *passing* and $progress(x)$ characterizing the progress of passing a signal to the target neuron with speed v and $x \in [0, 1]$.

To avoid the confusion of events/fluents of different neurons, we mark each fluent and each event by subscripts.

- $fires_{\langle i\rangle}$ expresses that neuron i fires;
- $energy_{\langle i\rangle}(x)$ expresses neuron i is of energy x, and similarly for *inhibited*, *uninhibited* and *raising*[6];
- $stimulus_{\langle i|j\rangle}$ expresses the event that neuron i's stimulant signal is received by neuron j, and similarly for *inhibition* and *passing*.

Formally, we have the following causal laws for all successive neurons i

(1) $Hol(energy_{\langle i\rangle}(x), t) \wedge x < d \wedge \neg Hol(raising_{\langle i\rangle}, t) \rightarrow Ini(stimulus_{\langle j|i\rangle}, raising_{\langle i\rangle}, t)$

(2) $Hol(energy_{\langle i\rangle}(x), t) \wedge x + u \leq d \rightarrow Tra(raising_{\langle i\rangle}, t, excited_{\langle i\rangle}, x + u)$

(3) $Hol(uninhibited_{\langle i\rangle}, t) \rightarrow Ini(inhibition_{\langle j|i\rangle}, inhibited_{\langle i\rangle}, t)$

(4) $Hol(uninhibited_{\langle i\rangle}, t) \rightarrow Ter(inhibition_{\langle j|i\rangle}, uninhibited_{\langle i\rangle}, t)$

(5) $Hol(energy_{\langle i\rangle}(d), t) \wedge Hol(raising_{\langle i\rangle}, t) \wedge Hol(uninhibited_{\langle i\rangle}, t) \rightarrow Hap(fires_{\langle i\rangle}, t)$

There are also some less general causal laws which apply only locally in Figure 1

- $\neg Hol(passing_{\langle 1|2\rangle}, t) \rightarrow Ini(fires_{\langle 1\rangle}, passing_{\langle 1|2\rangle}, t)$
- $\neg Hol(passing_{\langle 1|2\rangle}, t) \rightarrow Rel(fires_{\langle 1\rangle}, progress_{\langle 1|2\rangle}(), t)$
- $progress_{\langle 1|2\rangle}(x) \wedge x + vd' \leq 1 \rightarrow Tra(passing_{\langle 1|2\rangle}, t, progress_{\langle 1|2\rangle}(x + vd'), d')$
- $Hol(progress_{\langle 1|2\rangle}(1), t) \wedge Hol(passing_{\langle 1|2\rangle}, t) \rightarrow Hap(stimulus_{\langle 1|2\rangle}, t)$

For the initial states of passing signals, we have $\neg Inl(passing) \wedge Inl(progress(0))$. With the three causal laws above, it can be confirmed straightforwardly that the following hold

- $\{Hap(fires_{\langle 1\rangle}, 0), Ini(fires_{\langle 1\rangle}, passing_{\langle 1|2\rangle}, 0)\} \succ Hol(passing_{\langle 1|2\rangle}, t)$ for all $t \in (0, \frac{1}{v}]$
- $Hol(progress_{\langle 1|2\rangle}(0), 0) \succ Tra(passing_{\langle 1|2\rangle}, 0, progress_{\langle 1|2\rangle}(1), \frac{1}{v})$

[6] *inhibited* and *uninhibited* express the two exclusive states of neurons, and *uninhibited* doesn't represent absence, instead, it represents a particular state as *inhibited* does

- $Tra(passing_{\langle 1|2\rangle}, 0, progress_{\langle 1|2\rangle}(1), \frac{1}{v}) \succ Hol(progress_{\langle 1|2\rangle}(1), \frac{1}{v})$ and $Hol(passing_{\langle 1|2\rangle}, t) \succ Hol(progress_{\langle 1|2\rangle}(1), \frac{1}{v})$ for all $t \in (0, \frac{1}{v}]$
- $Hol(progress_{\langle 1|2\rangle}(1), \frac{1}{v}) \succ Hap(stimulus_{\langle 1|2\rangle}, \frac{1}{v})$

which collectively imply that

$$Hap(fires_{\langle 1\rangle}, 0) \dashrightarrow Hap(stimulus_{\langle 1|2\rangle}, \frac{1}{v})$$

Informally, the stimulus received by neuron 2 is productively caused by the firing of neuron 1.

For all successive neurons, i.e. 2 and 3 in Figure 1, it holds that $Inl(energy(0)) \wedge Inl(uninhibited)$. As for initial neurons, i.e. 1, 3 and 5, neither their states nor their events are of interest until they fire.

Maybe the causal laws above are not sufficient for a complete story about the neuron's raising of energy level, but they can meet the need of discussion here since what interests theorists are only the process beginning with receiving signals, ending with going into excited state and possibly also with firing and the target neuron's receiving the signal.

Continuing the analysis of the causal production of neuron 2's firing, causal laws (1) - (4) enable us to figure out that the following hold,

- $\{Hap(stimulus_{\langle 1|2\rangle}, 0), Ini(stimulus_{\langle 1|2\rangle}, raising_{\langle 2\rangle}, 0)\} \succ Hol(raising_{\langle 2\rangle}, t)$ for all $t \in (0, d]$
- $Hol(energy_{\langle 2\rangle}(0), 0) \succ Tra(raising_{\langle 2\rangle}, 0, excited_{\langle 2\rangle}, d)$
- $\{Hol(raising_{\langle 2\rangle}, t), Tra(raising_{\langle 2\rangle}, 0, energy_{\langle 2\rangle}(d), d)\} \succ Hol(energy_{\langle 2\rangle}(d), d)$ for all $t \in (0, d]$
- $\{Hol(energy_{\langle 2\rangle}(d), d), Hol(raising_{\langle 2\rangle}, d), Hol(uninhibited_{\langle 2\rangle}, d)\} \succ Hap(fires_{\langle 2\rangle}, d)$

which imply that

$$Hap(stimulus_{\langle 1|2\rangle}, 0) \dashrightarrow Hap(fires_{\langle 2\rangle}, d)$$

The initial time point 0 here is the time when the stimulant signal from neuron 1 is received by neuron 2, so when fixing the initial time point as above, we have

$$Hap(stimulus_{\langle 1|2\rangle}, \frac{1}{v}) \dashrightarrow Hap(fires_{\langle 2\rangle}, d + \frac{1}{v})$$

and thus

$$Hap(fires_{\langle 1\rangle}, 0) \dashrightarrow Hap(fires_{\langle 2\rangle}, d + \frac{1}{v})$$

Since the only path for productive causation between neuron 3's and neuron 2's (absence of) firings, if there is, necessarily involves $Hap(inhibition_{\langle 3|2\rangle}, t)$ and in the Figure 1 we have $\neg Hap(inhibition_{\langle 3|2\rangle}, t)$ for all t, there is no productive causation between neuron 3 and neuron 2 and thus not between neuron 4/5 and neuron 2 in the actual situation at issue.

4.2 Causal Contribution

Different from Hall's proposal, the causation other than causal production is named as *causal contribution* rather than dependence. The strategy of identification of causal contribution is similar to that of causal production. The main difference between causal contribution and causal production is that absence can also be relata of causation.

As a consequence, double prevention can be a contribution to a particular effect but may well be irrelevant to its production. The absence of interfering factor is expressed by atomic facts of the form $\neg Cli(t, f, t')$. $\neg Cli(t, f, t')$ connects the causal chain to those events and fluents protecting f from being terminated while $\neg Dec(t, f, t')$ lets us figure out those which prevent f from holding, so $\neg Cli(t, f, t')$ and $\neg Dec(t, f, t')$ are of importance to the identification of causal contributions even if those events and fluents determining them never participate in the causal production of f.

Definition 4.2. Given domains D_e of events, D_f of fluents, Dom of atomic formulas and AT of atomic facts, a root $\mathcal{R} = \langle F_P, L_C\rangle$ with the completion $Comp(\mathcal{R}) = \langle F_P, L_D\rangle$. Let $M = \{(\neg)p \mid p \in AT \wedge \mathcal{P} \models (\neg)p\}$, $FE^+ = \{(\neg)p \in M \mid p$ is of the form $Hap(e, t)$ or $Hol(f, t)\}$, the *causal contribution* $\curvearrowright$ is the intersection of

(a) $FE^+ \times FE^+$ and

(b) the transitive closure $TC(\triangleright)$ of $\triangleright$.

where $\triangleright \subsetneq M \times M$ is defined as follows. For $(\neg)p, (\neg)p' \in M$, we say $(\neg)p$ is a *contributory cause* of $(\neg)p'$ or $(\neg)p$ *causally contributes to* $(\neg)p'$ if $(\neg)p \curvearrowright (\neg)p'$.

(1) For $p(t, d) \in \{Hap(e, t), Ini(e, f, t), Ter(e, f, t), Tra(f, t, f', d), Rel(e, f, t), Fix(e, f, t)\}$ with definition

$$Def(p(t, d)) = \bigvee_i \sigma_i \leftrightarrow p(t, d)$$

let $p(t_0, d_0)$ be the atomic fact given by substituting constants t_0 and d_0 for the variables t and d in $p(t, d)$.

(1.1) We have

$$(\neg)Hol(f, t_1) \triangleright p(t_0, d_0)$$

if there is i such that $\mathcal{P} \models \sigma_i(t_0, d_0)$ and $\sigma_i(t_0, d_0) \models (\neg)Hol(f, t_1)$.

(1.2) It holds that

$$(\neg)Hol(f,t_0) \triangleright \neg p(t_0,d_0)$$

if $\mathcal{P} \models (\neg \bigvee_i \sigma_i(t_0,d_0)) \wedge (\neg)Hol(f,t_1)$ and $\sigma_i(t_0,d_0) \models \overline{(\neg)}Hol(f,t_1)$ for some i where

$$\overline{(\neg)}Hol(f,t_1) = \begin{cases} \neg Hol(f,t_1) & \text{if } (\neg)Hol(f,t_1) = Hol(f,t_1) \\ Hol(f,t_1) & \text{otherwise} \end{cases}$$

When a fluent is involved in the antecedent of a causal law, its absence can be (a) an enabling condition of a change, e.g. absence of other planes enables a plane to take off and land, or (b) a disabling condition of a change, e.g. the absence of oxygen can prevent wood from being on fire.

The sub-definitions of negative atomic facts are likely to provide more information about contributions than others do since for each positive atomic fact there is normally a unique causal chain ending with it while we have to preclude all possible causal chains leading to a positive atomic fact if its negation holds.

As an example, for $Def(Ini(e,f,t)) = \bigvee_i \sigma_i \leftrightarrow Ini(e,f,t)$, when $\neg Ini(e,f,t)$ holds, that is, no mode of enabling e to initiate f holds, any fluent or absence of fluent falsifying a σ_i contributes to the absence of the potential of e to initiate f, which is captured by clause (1.2) of Definition 4.2.

(2) For $Hol(f,t) \in Dom$ such that $Hol(f,t_0) \in M$,

(2.1) if $\mathcal{P} \models InlHol_f(t_0)$, then

$$\neg Cli(0,f,t_0) \triangleright Hol(f,t_0)$$

(2.2) if $\mathcal{P} \models IniHol_e(t_0)$ for some $Ini(e,f,t') \in Dom$, then pick the t_1 witnessing the existential quantifier in $IniHol_e(t)$, we have

$$\begin{aligned} Hap(e,t_1) &\triangleright Hol(f,t_0) \\ Ini(e,f,t_1) &\triangleright Hol(f,t_0) \\ \neg Cli(t_1,f,t_0) &\triangleright Hol(f,t_0) \end{aligned}$$

(2.3) if $\mathcal{P} \models TraHol_{f'}(t_0)$ for some $Tra(f',t',f,d) \in Dom$, then pick the t_1, d_0 witnessing the existential quantifiers in $TraHol_{f'}$, we have

$$Hol(f',t_2) \triangleright Hol(f,t_0)$$

for all t_2 such that $t_1 < t_2 < t_1 + d_0$, and

$$\begin{aligned} Tra(f',t_1,f,d_0) &\triangleright Hol(f,t_0) \\ \neg Cli(t_1+d_0,f,t_0) &\triangleright Hol(f,t_0) \end{aligned}$$

(3) For $Hol(f,t) \in Dom$ such that $\neg Hol(f,t_0) \in M$,

(3.1) if $\mathcal{P} \models \neg Hol(f,t)$ for all $t \in [0,t_0]$, then

$$\neg Dec(0,f,t_0) \triangleright \neg Hol(f,t_0)$$

(3.2) if there is $t_1 \in \mathbb{R}$, $Hap(e,t), Ter(e,f,t) \in Dom$ such that $\mathcal{P} \models Hap(e,t_1) \wedge Ter(e,f,t_1) \wedge \neg Hol(f,t_2)$ for all $t_2 \in (t_1,t_0]$, then we have

$$\begin{aligned} Hap(e,t_1) &\triangleright \neg Hol(f,t_0) \\ Ter(e,f,t_1) &\triangleright \neg Hol(f,t_0) \\ \neg Dec(t_1,f,t_0) &\triangleright \neg Hol(f,t_0) \end{aligned}$$

(3.3) if there is $t_1 \in \mathbb{R}$, $Hap(e,t), Rel(e,f,t) \in Dom$ such that

(a) $\mathcal{P} \models Hap(e,t_1) \wedge Rel(e,f,t_1)$
(b) there is a maximal $t_2 \in (t_1,t_0)$ such that $\mathcal{P} \models Hol(f,t_2)$ and
(c) for no $t_3 \in (t_1,t_2]$ there are $Hap(e',t), Fix(e',f,t) \in Dom$ with $\mathcal{P} \models Hap(e',t_3) \wedge Fix(e',f,t_3)$,

then we say

$$\begin{aligned} Hap(e,t_1) &\triangleright \neg Hol(f,t_0) \\ Rel(e,f,t_1) &\triangleright \neg Hol(f,t_0) \\ \neg Dec(t_2,f,t_0) &\triangleright \neg Hol(f,t_0) \end{aligned}$$

Moreover, for all $t' \in (t_1,t_2]$ and $Fix(e',f,t) \in Dom$, we have (i)

$$\neg Fix(e',f,t') \triangleright \neg Hol(f,t_0)$$

if $\mathcal{P} \models \neg Fix(e',f,t')$ and (ii)

$$\neg Hap(e',t') \triangleright \neg Hol(f,t_0)$$

if $\mathcal{P} \models Fix(e',f,t') \wedge \neg Hap(e',t')$.

The factors contributing to absence of fluents include not only those inhibitory ones but also those which terminate fluents. When it's asked why a fluent doesn't hold at a particular time, it can be reasonably answered that because it was terminated at some earlier time and failed to be initiated again since then.

When the absence of an event is thought of as a cause, it normally marks the absence of a change e.g. initiation, termination, release or fixing of a fluent. The absence of a fluent can also mark the absence of the initiation of a fluent if the former serves as an argument of Tra.

Timing is even more important to identification of causal contribution, for instance, in the inquiry on the absence of a fluent f at time t_0 (formally, $\neg Hol(f, t_0)$) resulted from a continuous change, it suffices to be the case that the continuous change is going on at the last instant t_2 when f holds, and in detail, the continuous change starts before t_2 and doesn't stop before t_2, as is represented by clause (3.3) of Definition 4.2. Again, those factors inhibiting the stop of the continuous change and those prevent f from being initiated between t_2 and t_0, also contribute to the absence of f at t_0.

(4) For $\neg Cli(t_0, f, t_1) \in M$

(4.1) for all $t_2 \in [t_0, t_1)$ and all $Ter(e, f, t) \in Dom$, it holds that (i)

$$\neg Ter(e, f, t_2) \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models \neg Ter(e, f, t_2)$, and (ii)

$$\neg Hap(e, t_2) \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models Ter(e, f, t_2) \wedge \neg Hap(e, t_2)$;

(4.2) for $Rel(e, f, t) \in Dom$,

(a) if there is $t_2 \in [0, t_0]$ such that $\mathcal{P} \models Hap(e, t_2) \wedge Fix(e, f, t_2)$ for some $Fix(e, f, t) \in Dom$, then pick the maximal $t_2 \in [0, t_0]$, it holds that

$$Hap(e, t_2) \triangleright \neg Cli(t_0, f, t_1)$$
$$Fix(e, f, t_2) \triangleright \neg Cli(t_0, f, t_1)$$

Moreover, for all $t' \in (t_2, t_1)$ and all $Rel(e', f, t) \in Dom$, we have (i)

$$\neg Rel(e', f, t') \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models \neg Rel(e', f, t')$, and (ii)

$$\neg Hap(e', t') \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models Rel(e', f, t') \wedge \neg Hap(e', t')$.

(b) if there is no $t_2 \in [0, t_0]$ such that $\mathcal{P} \models Hap(e, t_2) \wedge Fix(e, f, t_2)$, then for all $t' \in [0, t_1)$ and all $Rel(e', f, t) \in Dom$, we have (i)

$$\neg Rel(e', f, t') \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models \neg Rel(e', f, t')$, and (ii)

$$\neg Hap(e', t') \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models Rel(e', f, t') \wedge \neg Hap(e', t')$.

For $\neg Cli(t_0, f, t_1)$, if there is a continuous change started before t_0, then the stop of the last continuous change (at time t_2) before t_0 also contributes to the absence of interfering factors of f, and those preventing the start of continuous changes of f between t_2 and t_1 make contributions, too, as is represented by clause (4.2, a) of Definition 4.2. Clause (4.2, b) characterizes the case where no continuous change is stopped, which implies that no continuous change starts before t_0 under the assumption that $\neg Cli(t_0, f, t_1)$. This case is simpler since only those preventing the start of continuous changes of f between 0 and t_1 make contributions.

The case of $\neg Dec(t_0, f, t_1)$ is somehow simpler in which it just needs to be guaranteed that f can be initiated at no instant between t_0 and t_1.

(5) For $\neg Dec(t_0, f, t_1) \in M$

(5.1) for all $t_2 \in [t_0, t_1)$ and all $Ini(e, f, t) \in Dom$, it holds that (i)

$$\neg Ini(e, f, t_2) \triangleright \neg Cli(t_0, f, t_1)$$

if $\mathcal{P} \models \neg Ini(e, t_2)$, and (ii)

$$\neg Hap(e, t_2) \triangleright \neg Dec(t_0, f, t_1)$$

if $\mathcal{P} \models Ini(e, f, t_2) \wedge \neg Hap(e, t_2)$;

(5.2) for all $Tra(f', t, f, d) \in Dom$, all $t_2 \in [0, t_1)$, $d_0 \in \mathbb{R}^+$ such that $t_2 + d_0 \in (t_0, t_1]$, we have

(a)

$$\neg Tra(f', t_2, f, d_0) \triangleright \neg Dec(t_0, f, t_1)$$

if $\mathcal{P} \models \neg Tra(f', t_2, f, d_0)$;

(b)

$$\neg Hol(f', t') \triangleright \neg Dec(t_0, f, t_1)$$

for all $t' \in (t_2, t_2 + d)$ such that $\mathcal{P} \models \neg Hol(f', t')$ if $\mathcal{P} \models Tra(f', t_2, f, d_0)$.

When an event e with $Ini(e, f, t) \in Dom$ fails to initiate f at an instant t_2, there are two possibilities which are represented by (i) and (ii) of Clause 5.1. The first possibility is such that $Ini(e, f, t_2)$ doesn't hold, then the only the atomic fact $\neg Ini(e, f, t_2)$, which expresses that there is no potential or enabling condition for the initiation of f, contributes to the absence of initiation of f. In this case, even if e doesn't happen at t_2, formally, $\neg Hap(e, t_2)$, it won't be thought of as contributing to the absence of f's initiation since it's preempted by the absence of potential for the initiation of f. For example, it's weird to say that absence of spark contributes to the absence of fire in a bottle of pure water, since in pure water there is no enabling condition for the initiation of fire. Similarly for absence of termination, trajectory, release and fixing.

Applying the framework to the analysis of causal contribution in Figure 1, we have

- $\neg Cli(0, uninhibited_{\langle 2\rangle}, r+d) \triangleright Hol(uninhibited_{\langle 2\rangle}, r+d)$ where it holds that $Hap(stimulus_{\langle 1|2\rangle}, r)$
- $\neg Hap(inhibition_{\langle 3|2\rangle}, t) \triangleright \neg Cli(0, uninhibited_{\langle 2\rangle}, r+d)$ for all $t \in [0, r+d)$ since it holds that $Ter(uninhibited_{\langle 2\rangle}, inhibition_{\langle 3|2\rangle}, t)$
- $\neg Hap(fires_{\langle 3\rangle}, t-d') \curvearrowright \neg Hap(inhibition_{\langle 3|2\rangle}, t)$ for for all $t \in [d', r+d)$, where d' is the time for passing the signal between neuron 3 and 2
- $Hol(inhibited_{\langle 3\rangle}, t-d') \triangleright \neg Hap(fires_{\langle 3\rangle}, t-d')$ for $t \in (t'+d', r+d)$ where it holds that $Hap(inhibition_{\langle 4|3\rangle}, t')$
- $Hap(inhibition_{\langle 4|3\rangle}, t') \triangleright Hol(inhibited_{\langle 3\rangle}, t-d')$
- $Hap(fires_{\langle 4\rangle}, t'-d'') \curvearrowright Hap(inhibition_{\langle 4|3\rangle}, t')$ where d'' is the length of the time of passing signal between neuron 4 and 3.

It follows that

$$Hap(fires_{\langle 4\rangle}, t'-d'') \curvearrowright Hol(uninhibited_{\langle 2\rangle}, d)$$

Similar to the analysis of causal production, we have

$$Hol(uninhibited_{\langle 2\rangle}, r+d) \curvearrowright Hap(fires_{\langle 2\rangle}, r+d)$$

and thus it holds that

$$Hap(fires_{\langle 4\rangle}, t'-d'') \curvearrowright Hap(fires_{\langle 2\rangle}, r+d)$$

which expresses that neuron 4's firing contributes to that of neuron 2.

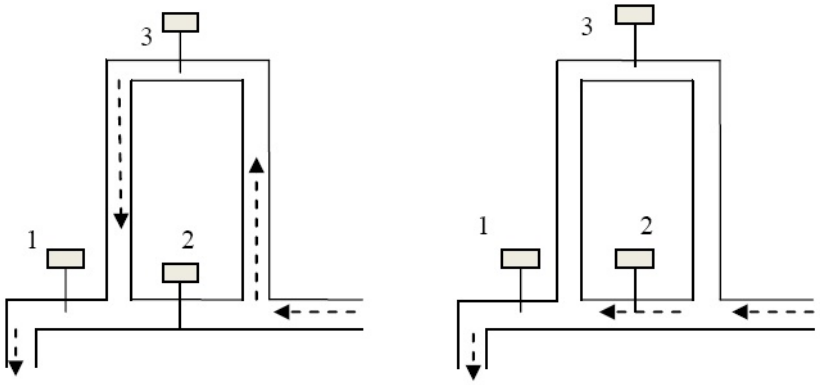
4.3 A constraint on the formulation of causal laws

There needs to be some further constraints on the formulation of causal laws involved in the mechanism of causal identification. An example is illustrated by the figure below, water flows out if and only if switch 1 and at least one of 2 and 3 are on.

When all three switches are on, water doesn't flow through switch 3 since it's much higher than switch 2. In this situation, suppose additionally that switch 2 and 3 are initially on while 1 is off, and one switches 1 on at time t_1, it's obvious that switch 3 doesn't produce or contribute to this initiation of water flow since water doesn't flow through switch 3 or force water to flow through switch 2. This intuition would contradict the formal identification of causation, if there is a causal law

$$Hol(on3, t) \wedge \neg Hol(on1, t) \rightarrow Ini(turnon1, flow, t)$$

since this causal law will lead to the claim that both $Hol(on3, t_1) \succ Ini(turnon1, flow, t_1)$ and $Hol(on3, t_1) \triangleright Ini(turnon1, flow, t_1)$ hold.



Some further constraint could be imposed on the causal laws formulated in the system, namely, each atomic fact of the form $(\neg)Hol(f,t)$ entailed by the antecedent of every causal law of the form $S(t) \rightarrow Ini(e,f,t)$ does feature in the causal history of the f's initiation whenever the antecedent $S(t)$ and $Hap(e,t)$ are satisfied. Similarly for Ter, Tra, Fix and Rel. So the causal law above doesn't meet this requirement; instead, we employ the following one

$$Hol(on3,t) \wedge \neg Hol(on1,t) \wedge \neg Hol(on2,t) \rightarrow Ini(turnon1, flow, t)$$

which implies that for switch 3 to contribute to the initiation of water flow, switch 2 should be off. Moreover, we have the causal law

$$\neg Hol(on1,t) \wedge Hol(on2,t) \rightarrow Ini(turnon1, flow, t)$$

which indicates that even when both switch 2 and switch 3 are on, there is no causation between the initiation of water flow by turning switch 1 on and switch 3's being on.

This constraint shouldn't be accused of begging the question of causal identification, since it's often the case that people needs to know not only the sufficient causes of a particular effect but also under what kinds of circumstances some particular causes will actually contribute to the effect. What we propose here is just to encode this type of further knowledge into the causal laws in our system.

What's more, each causal law represents a particular mode of effecting its consequent, so the antecedents of causal laws with the same consequent should be pairwise exclusive. Otherwise, if there is a situation satisfying the antecedents of two causal laws with the same consequent, we would have have to admit that the effect (token) at a particular time is caused in two different modes.

4.4 Linguistic expression of causation

Given the further constraint, in the situation where only switch 3 is initially on, a claim following from the formalizations is that switch 2's being off also produce and contribute

to the initiation of water flow if switch 1 is turned on. It's obvious that switch 2's being off does participated in the production of the particular mode of the initiation of water flow by forcing water to flow through switch 3. However, it's still somehow weird to say that water flows out *because* switch 2 is off. Switch 3's being on, but not switch 2's being off, is thought of as a cause of water flow since water flows through switch 3 rather than switch 2 though it's switch 2's being off that forces this particular route of water flow.

Another widely discussed example is about a persistent doctor and his unfortunate patient can be cited to illustrate the influence of pragmatic factors.

> The patient was dying and medical instruments showed that the patient couldn't remain alive after ten minutes. But the doctor didn't give up. He succeeded in saving the patient. However, because of serious disease, the patient died a few days later.

The patient's actual death lies at the end of its complicated causal chain where the doctor's effort in cure played an important role, since otherwise the patient's death would be earlier than it actually was. Nevertheless, no one is able to accuse the doctor of causing the patient's death, nor anyone could say that the patient died *because* the doctor saved him earlier without giving rise to strong objections from lots of listeners. The former perspective is typically about what Hall [3] calls production where there could be processes towards the contrary of the end of the causal chain and some events, such as the cure, which delay the occurrence of the result.

Moreover, though the birth of the patient is a necessary component in the causal chain of his illness and death, almost no one would like to say that his birth is a cause of his illness and death or he died because he was born.

A plausible explanation is that the causal expression in natural language is influenced by pragmatic factors, while the notion of causal contribution defined above is history-based, that is, $p \curvearrowright q$ says that p contributes to the particular history of the fact q. Among those contributories of a particular effect, some are such that people are often reluctant to say that the effect occurs because of these contributories, for various reasons one of which is that people often use expressions of causation in natural language to represent accountability ([14], p. 85). This point could be illustrated by the example about the doctor and his patient.

Linguistic expressions of causation is not the focus here as long as we notice the difference between causal production/contribution and linguistic expressions of causation, since the aim is the mechanism of identification of causation which should be exhaustive in order to ground the semantics of counterfactuals.

5 The semantics of counterfactuals

5.1 The semantics

The definitions of causal production $\dashrightarrow$ and causal contribution $\curvearrowright$ can be defined more generally over each fixed model of $\mathcal{P}$ when $\mathcal{P}$ isn't sufficient for determining a unique full model for those atomic facts of AT.

The independence of particular facts are determined by causal contribution rather than causal production. As can be seen in Definition 4.1 and 4.2, causal contribution is an extension of causal production if regardless of pragmatic considerations concerning linguistic representations of causation. The difference between the two notions causation is crucial for counterfactual reasoning. For example, we can have $\neg Cli(t_1, f, t_2) \triangleright Hol(f, t_2)$ if we have $Hol(f, t)$ for all $t \in [t_1, t_2]$ but never $\neg Cli(t_1, f, t_2) \succ Hol(f, t_2)$. If it has been the case that $Cli(t_1, f, t_2)$, then we have to abandon $Hol(f, t_2)$ first though it may be forced by backup causes later.

Definition 5.1. Given the domains and a root $\mathcal{R} = \langle F_P, L_C \rangle$ with completion $Comp(\mathcal{R}) = \langle F_P, L_D \rangle$, the family IC of integrity constraints which contains (1a), (1b), (2a) and (2b), and a model $\mathcal{M}$ of $\mathcal{P} = F_P \cup L_D \cup IC$ for AT. Let $M = \{(\neg)p \mid p \in AT \wedge \mathcal{M} \models (\neg)p\}$. An atomic fact $(\neg)p \in M$ is *causally independent* if (1) p is of the form $Inl(f)$ or (2) p is of the form $Hap(e, t)$ and there is no $(\neg)q \in M$ with $(\neg)q \curvearrowright (\neg)p$.

The *basis* $B_{\mathcal{M}}$ of $\mathcal{M}$ consists of exactly those causally independent elements of M.

It could be concluded that an atomic fact $(\neg)p \in M$ is causally independent if (i) p is of them form $Inl(f)$ or (ii) p is of the form $Hap(e, t)$ and there is no causal law with the consequent $Hap(e, t)$, since $\mathcal{M}$ is a model of $\mathcal{P}$. Intuitively, $B_{\mathcal{M}}$ consists of the causally independent *particular* facts.

As has been pointed out, atomic facts $(\neg)q$ of the form $(\neg)Ini(e, f, t)$, $(\neg)Ter(e, f, t)$, $(\neg)Tra(f, t, f', d)$, $(\neg)Rel(e, f, t)$, $(\neg)Fix(e, f, t)$ express law-like facts and thus should have their causal laws (possibly with empty antecedent) in L_C. If the causal law defining q has empty antecedent, then q holds universally as a law throughout the time in question and is encoded in L_C. If the causal has non-empty antecedent, then $(\neg)q \in \mathcal{M}$ must be dependent on other atomic facts according to Definition 4.2. In both cases, $(\neg)q$ doesn't qualify as an element of $B_{\mathcal{M}}$.

Atomic facts $(\neg)q$ of the form $(\neg)Cli(t, f, t')$ or $(\neg)Dec(t, f, t')$ depend on f-relevant events and fluents (if there are), i.e. those which can initiate, release, terminate or fix f, so these $(\neg)q$ is in no means independent. If there is no event taken into account which can terminate or release f, formally, Dom contains no $Ter(e, f, t)$ or $Rel(e, f, t)$, then the definition of $Cli(t, f, t')$ would be

$$\bigvee \emptyset \leftrightarrow Cli(t, f, t')$$

and thus $\neg Cli(t, f, t')$ holds universally like a law, which is encoded as the absence of $Ter(e, f, t)$ and $Rel(e, f, t)$ in Dom and doesn't need to be represented as (independent) particular facts.

Atomic facts $(\neg)q$ of the form $(\neg)Hol(f, t)$ aren't independent facts. Considering the principle of inertia, there are two possibilities. (i) If it's the case that $(\neg)Hol(f, t)$ holds since time 0, then it depends on the absence of f-relevant events and fluents. (ii) If q is initiated, terminated or released, $(\neg)q$ isn't independent, either.

Definition 5.2. Given the domains, a root $\mathcal{R} = \langle F_P, L_C \rangle$, the family IC of integrity constraints which contains (1a), (1b), (2a) and (2b), and a model $\mathcal{M}$ of $\mathcal{P} = F_P \cup L_D \cup IC$, a subset F of AT^+ is *R&IC-consistent* if there is no $p \in AT$ such that $F \cup L_D \cup IC \models p \wedge \neg p$[7].

A *R&IC*-consistent subset F of AT^+ is *$\mathcal{R}$-basic* if for all $(\neg)p \in F$, either (i) p is of the form $Inl(f)$ or (ii) p is of the form $Hap(e, t)$ and there is no causal law with the consequent $Hap(e, t)$. The relative similarity relation $<_{\mathcal{M}}$ over $\mathcal{R}$-basic sets is defined as follows.

Given two $\mathcal{R}$-basic sets F and F', F is *more similar to* $\mathcal{M}$ than F' is (formally, $F <_{\mathcal{M}} F'$) iff the following holds

$$(B_{\mathcal{M}} - F) \cup (F - B_{\mathcal{M}}) \subsetneq (B_{\mathcal{M}} - F') \cup (F' - B_{\mathcal{M}})$$

Definition 5.3. Given the domains, a root and a model as above, a counterfactual $\varphi \hookrightarrow \psi$ holds at $\mathcal{M}$ (notation: $\mathcal{M} \models \varphi \hookrightarrow \psi$) iff for all $\mathcal{R}$-basic F such that (i) $F \cup L_D \cup IC \models \varphi$ and (ii) there is no $\mathcal{R}$-basic F' with $F' \cup L_D \cup IC \models \varphi$ and $F' <_{\mathcal{M}} F$, it holds that $F \cup L_D \cup IC \models \psi$.

As can be seen in the definitions, with fixed domains, there is no violation of causal laws or *definitions* in counterfactual reasoning. This stipulation is developed from the widely accepted principle proposed by Lewis [8]

> 'It is of the first importance to avoid big, widespread, diverse violations of law.'

There are violations of laws in conditional or causal reasoning, but violations are usually accompanied with acquisition or retrieval of more interfering factors and thus extension of domains and causal laws. The intimate connection between the likelihood of violating laws and the retrieval of disabling conditions has been justified experimentally by Cummins [2]. When the factors involved in a causal structure are fixed and the causal laws at issue are formulated in complete forms, that is, in the forms where all disabling conditions are taken into account, causal laws shouldn't be violated.

As an example, apply this definition to Figure 1. The mechanism of energy level's raising applies to successive neurons, i.e. neuron 2 and 3 only, since the situation at issue

[7] $AT^+ = \{(\neg)p \mid p \in AT\}$

doesn't contain any information about the signals that neuron 1, 4 or 5 receives and thus it doesn't make sense to talk about their mechanisms of energy raising. Therefore, there is no causal law with the consequent $Hap(fires_{\langle i \rangle}, t)$ for $i \in \{1, 4, 5\}$. The independent facts in the situation include

- neuron 2 and 3 are initially uninhibited;
- neuron 2 and 3 are initially of energy 0;
- neuron 1, 4 and 5 fire (once respectively).

The time points at which the neurons fire are in particular relations but these relations are often neglected in discussions

(i) neuron 4's firing should be early enough so that neuron 3 is inhibited before going into excited state;

(ii) neuron 5's firing should be early enough in order to stimulate neuron 3 to inhibit neuron 2 before it goes into excited state if neuron 4 doesn't fire.

To judge the truth value of the counterfactual *if neuron 4 hadn't fired, neither would neuron 2 have*, we can keep all independent facts listed above except that neuron 4 fires (at some particular time t_0), and add the fact that $\neg Hap(fires_{\langle 4 \rangle}, t_0)$. This modification of the basis is necessary; otherwise, it can't be the case that neuron 4 hadn't fired. It follows that it's the unique minimal modification to satisfy the antecedent *if neuron 4 hadn't fired*. Then it holds that neuron 2 wouldn't have fired, as can be checked.

5.2 Allowing for vagueness

An example about King Ludwig has been widely discussed by theorists working on semantics of counterfactuals (e.g. [5] and [17]).

> King Ludwig often spends his holiday at his castle. Whenever the flag is up and the lights are on, King Ludwig is in the castle. Currently, the flag is down, the lights are on and the King is not in the castle.

Kratzer [5] attempts to defend the truth of the following counterfactual

(7) If the flag were up, then the king would be in the castle.

for which she has to eliminate the possibility that the lights are off and the king is away under the counterfactual assumption that the flag were up while this possibility is kept by Veltman [17].

What is agreed by all parties is, it's a non-accidental or lawlike generalization that whenever the flag is up and the lights are on, King Ludwig is in the castle. In the theory developed above, it could be represented as an integrity constraint.

It might be proposed that this lawlike generalization should be interpreted causally and decomposed into a group of causal laws. But there is almost no information about the causal mechanism of this lawlike generalization provided in the story. Any further detail added by analysts is likely to restrict the story to a particular version among the numerous possible ones, for instance, it's hard to specify if the states of the flag and the lights are determined by the king's activity, or the converse, or neither of them holds. Therefore, the most plausible way to represent the story at issue is such that the states of the three parties are causally independent of each other.

The analysis under event calculus can contain the following primitives.

- $D_e = \{raise, lower, turnon, turnoff, return, leave\}$
- $D_f = \{up, down, on, off, in, away\}$
- $Dom = \{Hap(e,t) \mid e \in D_e\} \cup \{Hol(f,t) \mid f \in D_f\} \cup \{Ini(raise, up, t)$, $Ter(raise, down, t)$, $Ini(lower, down, t)$, $Ter(lower, up, t)$, $Ini(turnon, on, t)$, $Ter(turnon, off, t)$, $Ini(turnoff, off, t)$, $Ter(turnoff, on, t)$, $Ini(return, in, t)$, $Ter(return, away, t)$, $Ini(leave, away, t)$, $Ter(leave, in, t)\}$

Assume that the period at issue is $[0, t_0]$ where t_0 is the current time, so

$$AT = \{Hap(e,t) \mid e \in D_e \wedge t \in [0,t_0]\} \cup \{Hol(f,t) \mid f \in D_f \wedge t \in [0,t_0]\} \cup \{Inl(f) \mid f \in D_f\}$$

The root $\mathcal{R} = \langle F_P, L_C \rangle$ where $F_P = \{\neg Hap(e,t) \mid e \in D_e, t \in [0,t_0]\} \cup \{Inl(on)$, $\neg Inl(off)$, $Inl(down)$, $\neg Inl(up)$, $Inl(away)$, $\neg Inl(in)\}$ and L_C contains the following causal laws

- $Hol(down, t) \rightarrow Ini(raise, up, t)$
- $Hol(down, t) \rightarrow Ter(raise, down, t)$
- $Hol(up, t) \rightarrow Ini(lower, down, t)$
- $Hol(up, t) \rightarrow Ter(lower, up, t)$
- $Hol(off, t) \rightarrow Ini(turnon, on, t)$
- $Hol(off, t) \rightarrow Ter(turnon, off, t)$
- $Hol(on, t) \rightarrow Ini(turnoff, off, t)$
- $Hol(on, t) \rightarrow Ter(turnoff, on, t)$
- $Hol(in, t) \rightarrow Ini(leave, away, t)$
- $Hol(in, t) \rightarrow Ter(leave, in, t)$
- $Hol(away, t) \rightarrow Ini(return, in, t)$
- $Hol(away, t) \rightarrow Ter(return, away, t)$

The definitions of atomic formulas follows immediately and thus so does $Comp(\mathcal{R})$.

The family IC consists of the following integrity constraints apart from (1a), (1b), (2a) and (2b)

(i) $Hol(up,t) \wedge Hol(on,t) \rightarrow Hol(in,t)$

(ii) $\neg(Hol(up,t) \wedge Hol(down,t))$

(iii) $\neg(Hol(on,t) \wedge Hol(off,t))$

(iv) $\neg(Hol(in,t) \wedge Hol(away,t))$

Kratzer wants to defend the truth of the counterfactual $Hol(up,t_0) \hookrightarrow Hol(in,t_0)$. The model $\mathcal{M}$ representing the situation at issue is such that no change occurs, and $Hol(on,t)$, $Hol(down,t)$ and $Hol(away,t)$ are true for all $t \in [0,t_0]$. By definitions, the basis B equals F_P. Pick $t_1, t_2 \in (0,t_0)$ with $t_1 < t_2$, let

$$F = (F_P - \{\neg Hap(turnoff,t_1), \neg Hap(raise,t_2)\}) \cup \{Hap(turnoff,t_1), Hap(raise,t_2)\}$$

It could be seen that F is $\mathcal{R}$-basic, $F \cup L_D \cup IC \models Hol(up,t_0)$, and there is no $\mathcal{R}$-basic F' such that $F' \cup L_D \cup IC \models Hol(up,t_0)$ and $F' <_{\mathcal{M}} F$. Since $F \cup L_D \cup IC \not\models Hol(in,t_0)$, it follows that $F \cup L_D \cup IC \not\models Hol(up,t_0) \hookrightarrow Hol(in,t_0)$.

This result accords with that gained in [17] and contradicts what is proposed by Kratzer [5]. If the analysis is made merely based on the original version of the story, then it seems that no convincing argument or intuition has been provided for the truth of (7) which requires that the state of lights wouldn't change under the counterfactual assumption that the flag were up.

What is likely to make us a bit unease in the formal analysis of the story about King Ludwig is, there could be some deeper causal relations between the states of the lights and the flag and the activity of King Ludwig which haven't been able to be specified, e.g. the former are somehow determined by the later. However, the current story only enables us to represent the three parties as almost independent of each other except that they conform to the integrity constraint (i).

It's logically possible that further details, if supplemented, would force the truth of (7). But if fixing merely the original version of the story, we have no way to preclude the possibility that the lights are off and the king remains away. Given partial information, the semantics should allows for the vagueness consisting in the situation rather than eliminate it artificially.

The same case holds for the counterfactual assumption about kangaroos. Since there is no information about the mechanism of kangaroos's evolution in the past millions of years, it can't be determined that if they would have developed some other way to keep their body balance if they hadn't got tails.

5.3 Epistemic reading of counterfactuals

Given the somehow complicated formalization above, there still seem to be a gap between its formal prediction and people's intuitions about some particular examples

which is worth big attention of anyone who aims at a better understanding of counterfactuals. Lots of theorists argue that there are two readings of conditionals, namely, epistemic reading and ontic reading. Schulz interpret the two readings as follows.

> 'The epistemic reading is based on belief revision. It is used for conditionals that make statements about what one would conclude upon learning that the antecedent is true. It reasons about what you would believe, if you learned - hypothetically - that the antecedent is true.
>
> ... ontic reading of conditionals ... is applied if the conditional is interpreted as describing the consequences for the course of history it would have, if the antecedent is true.' ([11], p. 127)

The ontic reading of conditionals is analyzed based on the principle of similarity with respect to causal dependence, as is discussed above, while the epistemic reading, according to Schulz's proposal, deserves a quite different treatment.

A widely discussed example is about a murder ([17], p. 11; [11], p. 129).

> The duchess was murdered, and the task of finding out the murder has been assigned to a detective. Currently, the gardener and the butler are the only ones left as suspects. Then the detective believes that
>
> (8) If the butler didn't kill the duchess, the gardener did.
>
> Suppose that after a few days of investigation, the detective found sufficient evidence showing that the butler committed the murder while the gardener is innocent. At that moment, would the detective believe the following sentence?
>
> (9) If the butler hadn't killed the duchess, the gardener would have.

The domains are as follows

$$\begin{aligned} D_e &= \{butl, gard\} \\ D_f &= \{alive, dead\} \\ Dom &= \{Inl(alive), Inl(dead), Hol(alive,t), Hol(dead,t), Ini(butl,dead,t), \\ &\quad Ter(butl,alive,t), Ini(gard,dead,t), Ter(butl,alive,t)\} \end{aligned}$$

where *butl* is short for *the butler killed the duchess*, *gard* for *the gardener killed the duchess*, *alive* for *the duchess was alive* and *dead* for *the duchess was dead*.

The family L_C of causal causal laws at issue consists of the following ones

- $Hol(alive,t) \rightarrow Ter(butl,alive,t)$
- $Hol(alive,t) \rightarrow Ini(butl,dead,t)$
- $Hol(alive,t) \rightarrow Ter(gard,alive,t)$

- $Hol(alive, t) \rightarrow Ini(gard, dead, t)$

Without loss of generality, assume that the murder happened at t_1 and currently the time is t_0. Then the basis

$$B = \{Inl(alive), \neg Inl(dead), Hap(butl, t_1)\} \cup \{\neg Hap(butl, t) \mid t \in [0, t_1) \cup (t_1, t_0]\} \cup \{\neg Hap(gard, t) \mid t \in [0, t_0]\}$$

To accommodate the antecedent *if the butler hadn't killed the duchess*, a necessary operation on the basis is abandoning the atomic fact $Hap(butl, t_1)$ and adding $\neg Hap(butl, t_1)$, formally, let

$$B' = (B - \{Hap(butl, t_1)\}) \cup \{\neg Hap(butl, t_1)\}$$

and this operation is sufficient for making the counterfactual assumption. Since for no $t \in [0, t_0]$ it holds that $B' \cup L_D \models Hap(gard, t)$, (9) is false.

To accept (9), it would have to be supposed that there is a conspiracy according to which the gardener serves as a backup of the butler, that is, the gardener would kill the duchess in case the butler failed to make it. But the fact is, there is no such conspiracy or it hasn't been discovered by the detective if there is, so nothing could support him to believe that the gardener would have killed the duchess in case the butler hadn't.

However, it might still be insisted that there is a reading under which (9) is true, i.e. epistemic reading. The formalizations in [17] and [11] are given in terms of propositional language. Let *butl* and *gard* be short for the same expressions as above and *duch* be short for *the duchess was alive*. The general law at issue is $butl \vee gard \rightarrow duch$. After observing that the duchess was killed and restricting the range of suspects to the butler and the gardener, the butler believes that $duch \wedge (butl \vee gard)$. Further evidence enables him to believe that $duch \wedge butl$. Making the counterfactual assumption that the butler hadn't killed the duchess (formally, $\neg butl$), the belief in *gard* is forced by the assumption $\neg butl$ plus that $butl \vee gard$, in other words, while the belief that *butl* is abandoned, it remains believed that $butl \vee gard$.

Why is the belief that $butl \vee gard$ kept under the counterfactual assumption that $\neg butl$? There are different accounts for this keep. Veltman proposes that this keep is actually some implicit reference to previous epistemic state ([17], p. 12). Then it seems that the subject is making some correction rather than counterfactual assumption. When making a correction, the subject only needs to abandon those beliefs which are definitely falsified while others are kept. In the example at issue, even if there is convincing evidence against the belief that the butler killed the duchess, it might not falsify the belief that the butler and the gardener are the only suspects.

A consequence of this interpretation of epistemic reading is, epistemic reading is rather hard to communicate since it depends on the epistemic state of the subject as well as the process of update of epistemic states ([11], p. 129). Then, even for those who share the same epistemic states, they could disagree on the truth value the same counterfactual under epistemic reading.

Schulz [11] claims that Veltman misunderstands epistemic reading of conditionals since in her opinion, the reference to previous epistemic states isn't an essential property of epistemic reading. To illustrate this point, she comes up with a slightly modified version of the story about murder ([11], p. 130).

> The duchess was murdered last night. The detective is supposed to find out the murder. Finger prints of the butler are found all over the crime scene, so the detective interrogates the butler and the butler confesses the murder, which makes the detective believe that the butler killed the duchess while the gardener has nothing to do with the criminal. Later, the lab reports that no lock of the house is broken. Besides the duchess, only the butler and the gardener have the keys to the house. In this situation, would the detective believe that (9) holds?

Schulz thinks that the detective would believe (9) in the modified version of story though there is no previous state where (8) holds. Nevertheless, someone else could still have the intuition that the detective wouldn't believe (9). It doesn't help much if the two sides are engaged in a disputation merely by appealing to intuitions they don't share.

Schulz proposes her formalization of epistemic reading in terms of belief state under the framework of premise semantics. Mainly, a belief state S is a pair $\langle \mathcal{B}, U \rangle$ where $\mathcal{B}$ is a finite set of sentences of propositional language which are built from propositional letters, negations and conjunctions, and U is a set of possible worlds such that $\mathcal{B}$ is satisfiable in U. A possible world is identified with a function whose domain is the family of propositional letters in question and whose range is $\{0, 1\}$. $\mathcal{B}$ is called the *basis* of S, which is stipulated to contain exactly those sentences for which the subject has gained independent external evidence ([11], p. 132). U is called the *universe* and contains those possible worlds satisfying the general laws acquired by the subject; alternatively, it can be said that the general laws are encoded in the universe of belief state.

Given a counterfactual $\varphi \hookrightarrow \psi$, a subject in belief state S believes that $\varphi \hookrightarrow \psi$ if and only if he believes ψ when he (hypothetically) learns φ. The hypothetical belief state after the subject learns that φ is determined by the following criteria

- no general law is violated;
- the subject believes that φ;
- the belief revision is minimal.

Specifically, the hypothetical belief state is represented by the $<_S$-minimal elements of $[\![\varphi]\!] \cap U$ where $<_S$ is the relative similarity relation defined as follows: for $w, w' \in [\![\varphi]\!] \cap U$, $w < w'$ iff for all $\chi \in \mathcal{B}$, if χ holds at w' then also at w, and the converse *doesn't* hold.

Applying this semantics to the story about murder, the belief state $S = \langle \mathcal{B}, U \rangle$ where $\mathcal{B} = \{duch, butl \vee gard, butl\}$ and U consists of those worlds satisfying $butl \vee gard \rightarrow duch$. Different from the version of premise semantics proposed by Veltman [17], the basis $\mathcal{B}$ is a set of sentences of propositional language which don't have to be logically independent of each other, e.g. $butl \vee gard$ and $butl$. As has been stated, the elements of basis are those sentences for which the subject has gained evidence. Since the detective have evidence for $butl \vee gard$ and $butl$ respectively, both of them belong to $\mathcal{B}$.

To make the counterfactual assumption $\neg butl$, the detective would have to abandon the sentence $butl$ while $duch$ and $butl \vee gard$ are kept according to the minimality of belief revision. Considering that $\neg butl$, it could be concluded that $gard$ and thus (9) holds under epistemic reading.The key point in this formal analysis is that $butl \vee gard$ is kept in the subject's making counterfactual assumption.

Given Schulz's arguments and explanations for the semantics, this keep may well be problematic. The elements of $\mathcal{B}$ is justified by their evidence respectively. Since the subject of belief states is presupposed to be rational, without which it would be extremely difficulty to give a formal representation of his interpretation and use of language, he should also suppose that (i) there is some convincing hypothetical evidence for which he would hypothetically believe that $\neg butl$ when making the counterfactual assumption, e.g. the butler has never accessed the crime scene or he was sleeping when the duchess was murdered, and additionally, (ii) the evidence justifying the detective's belief that $butl$, namely, finger prints and confession of the butler, can't remain in the counterfactual worlds. But these revisions are likely to preclude the happening of the murder, so the duchess may be alive.

Although the detective could imagine that the butler employed some tricks by which he killed the duchess without entering her room, which would probably serve as evidence for his innocence and make the detective believe that the gardener is left as the only suspect. But it will involve the extension of the situation at issue and the fact that the butler didn't employ the trick in the actual world. Then, the evidence for the butler's guilt, e.g. his finger prints and confession, depends on the event that he killed the duchess and the fact that he didn't employ any effective trick. As a consequence, when the detective supposes counterfactually that there is no evidence for the butler's guilt and instead there is evidence for his innocence, it might be equally probable either that he employed some trick or that the event of kill didn't happen.

More generally, there is often some dependence (most probably, causal dependence) between evidence justifying different beliefs. However, Schulz seems to implicitly assume that evidence for different sentences in $\mathcal{B}$ is independent of each other, or the subject of belief states doesn't need to consider the evidence for the hypothetical belief when making counterfactual assumption. Either of the two possible assumptions underlying Schulz's semantics of epistemic reading is likely to deprive belief's basis in reality, which would bring belief into the danger of becoming illusion.

Therefore, even if there does exist an epistemic reading of counterfactuals, it seems that its mechanism of meaning is still rather vague which makes it hard to communicate.

5.4 Situations with contingent factors

There is another controversial example about Kennedy's death which is similar to the duchess example.

> There was a conspiracy against Kennedy. According to this conspiracy, Oswald was the first one to shoot Kennedy. By accident, Oswald succeeded in killing Kennedy. To guarantee that Kennedy couldn't survive, there was another assassin Aswald who serves as a backup of Oswald. Aswald is such a skillful assassin that he never fails in his tasks. But considering that anyone who killed the president can hardly escape and the boss doesn't want to sacrifice him if it's unnecessary, Aswald is only assigned the task of shooting Kennedy in case Oswald missed the target. Then the following sentence holds
>
> (10) If Oswald hadn't killed Kennedy, Aswald would have.

At the first glance, the truth of (10) is dubious. Oswald's shot is resulted from the conspiracy against Kennedy, though only by accident, which implies that this shot isn't independent while it's independent that there is a conspiracy. It follows that a minimal revision of independent facts is abandoning the fact that there is a conspiracy and adding its negation. With this revision, neither assassin would have killed Kennedy and thus (10) doesn't hold.

Schulz [11] avoids the problem by precluding the proposition about the existence of conspiracy from the set of atomic propositions and thus the conspiracy is taken for granted, then it is a causal law that Aswald would fire in case Oswald missed the target. Nevertheless, it hasn't been explained well why the conspiracy could be taken as granted. Since the story, when told in [11], also mentions the existence of conspiracy which is obviously crucial to the death of Kennedy, the soundness of analysis would be unjustified if the crucial fact is taken as granted without being well motivated.

The point is likely to be, Oswald killed Kennedy *by accident*, that is, although Oswald's kill is caused by the conspiracy against Kennedy, there is some contingent and usually also complicated factor (denoted by X) determining the success of this assassination. Lots of factors are widely known to be important to snipe, e.g. speed and direction of wind, humidity level.

This complicated causal structure of relevant factors are usually far beyond the capture of most people, so this structure is usually summarized as a collective factor, which is named as X here. X can't be neglected or taken as fixed, since it can make a difference between the actual world and those counterfactual worlds, and particularly, it underlies the contingency of Oswald's success.

In event calculus, X can be represented as a special fluent such that X doesn't belong to D_f, which implies that X doesn't need to satisfy the principle of inertia since the axioms of event calculus, as is stated above, are only given respect to the domains D_e and D_f; instead, if X holds at t, i.e. the truth of $Hol(X,t)$, is absolutely independent and contingent. Formally, no atomic formula of the form $Inl(X)$, $Ini(e,X,t)$, $Ter(e,X,t)$, $Tra(f,t,X,d)$, $Rel(e,X,t)$ or $Fix(e,X,d)$ belongs to Dom and $Def(X)$ is always undefined. A causal law involving X is normally of the form

$$S(t) \wedge Hol(X,t) \rightarrow p(t)$$

Moreover, X can influence the result of Oswald's shot in a subtle way, for instance, a little difference in wind direction can lead to the difference between Oswald's success and failure. More generally, under the framework of event calculus, the collective factor X contained in the antecedent of a causal law can usually be assumed to have the power to influence the truth value of the consequent in a subtle way. This assumption can be justified as follows. Suppose the subject has learned that X cannot influence $p(t)$ in a subtle way, then this acquisition would normally requires that the subject has learned a specific range within which any change of X doesn't affect the truth of $p(t)$. There are two possibilities when $S(t)$ and $p(t)$ are fixed.

(i) Within this range, $p(t)$ is necessarily true, then a causal law without X should be added

$$S(t) \wedge S'(t) \rightarrow p(t)$$

such that $S'(t)$ represents the truth of some fluents involved in X under which no change of X can affect the truth of $p(t)$. Moreover, the causal law involving X is modified as

$$S(t) \wedge \neg S'(t) \wedge Hol(X,t) \rightarrow p(t)$$

that is, only under the condition that $S(t) \wedge \neg S'(t)$, the truth of $p(t)$ bears the contingency brought by X.

(ii) within this range, $p(t)$ is necessarily false, then the causal law involving X is modified as

$$S(t) \wedge \neg S'(t) \wedge Hol(X) \rightarrow p(t)$$

Take a modified version of the example about Jones's hat as an illustration ([17], [11]).

> Jones has such a disposition. At time t_0 of every day, he decides if he wears his hat. The bad weather at t_0 invariably induces him to wear his hat while fine weather doesn't influences his decision, that is, whether he puts on his hat or not is random.

In this situation, X probably represents the collective effect of Jones's mood, tightness of his schedule (whether he is in a hurry), Jones's willingness to put his hat on, etc. It seems that he puts on his hat randomly when the weather is good because he makes different decisions even when almost no notable change is observed. Instead, a tiny difference in his mood or willingness is likely to be sufficient for his making a different decision. Then the causal law acquired by the subject is

- $Hol(bad, t) \wedge t = t_0 \rightarrow Hap(puton, t)$
- $Hol(fine, t) \wedge Hol(X, t) \wedge t = t_0 \rightarrow Hap(puton, t)$

Suppose later the subject finds that when the weather is fine, Jones's being happy also invariably makes him wear his hat. Then the second causal law is replaced by two new ones

- $Hol(fine, t) \wedge Hol(happy, t) \wedge t = t_0 \rightarrow Hap(puton, t)$
- $Hol(fine, t) \wedge \neg Hol(happy, t) \wedge Hol(X, t) \wedge t = t_0 \rightarrow Hpa(puton, t)$

Given that X influences $p(t)$ in a subtle way, the necessary change accompanied with the hypothetical revision of the truth of $Hol(X, t)$ is extremely small, though often without specific representation in the mind of the subject. Then, compared with other standard atomic facts in question, namely, those contained in AT^+, $Hol(X, t)$ is assumed to be of a notably *lower* weight in the measure of relative similarity between possible worlds.

To accommodate the collective factor X in the formalizations developed above which serves as the source of contingency, the relation of relative similarity is to be re-defined. Let

$$AT_X = AT \cup \{Hol(X, t) \mid t \in T\}$$
$$AT_X^+ = \{(\neg)p \mid p \in AT_X\}$$

where T is the temporal interval during which the truth of $Hol(X, t)$ is of importance. There can be more than one X in question since several atomic facts can be determined by contingent factors which are independent of each other. As an illustration, we just discuss the case of a single X.

Definition 5.4. Given the domains and a root $\mathcal{R} = \langle F_P, L_C \rangle$ with the completion $Comp(\mathcal{R}) = \langle F_P, L_D \rangle$, the family IC of integrity constraints which contains (1a), (1b), (2a) and (2b), and a model $\mathcal{M}$ of $\mathcal{P} = F_P \cup L_D \cup IC$ for AT. Let $M = \{(\neg)p \in AT_X^+ \mid \mathcal{M} \models (\neg)p\}$. An atomic fact $(\neg)p \in M$ is *causally independent* if one of the following holds

(1) p is of the form $Inl(f)$

(2) p is of the form $Hap(e, t)$ and there is no $(\neg)q \in M$ with $(\neg)q \curvearrowright (\neg)p$;

(3) p is of the form $Hol(X, t)$.

The *basis* $B_{\mathcal{M}}$ of $\mathcal{M}$ consists of exactly those causally independent elements of M.

Definition 5.5. Given the domains, a root $\mathcal{R} = \langle F_P, L_C \rangle$, the family IC of integrity constraints which contains (1a), (1b), (2a) and (2b), and a model $\mathcal{M}$ of $\mathcal{P} = F_P \cup L_D \cup IC$, a subset F of AT_X^+ is *R&IC-consistent* if there is no $p \in AT_X$ such that $F \cup L_D \cup IC \models p \wedge \neg p$.

A $R\&IC$-consistent subset F of AT_X^+ is *$\mathcal{R}$-basic* if for all $(\neg)p \in F$, one of the following holds:

(i) p is of the form $Inl(f)$;

(ii) p is of the form $Hap(e, t)$ and there is no causal law with the consequent $Hap(e, t)$;

(iii) p is of the form $Hol(X, t)$.

The relative similarity relation $<_{\mathcal{M}}$ over $\mathcal{R}$-basic sets is defined as follows: given two $\mathcal{R}$-basic sets F and F', $F <_{\mathcal{M}} F'$ iff one of the following holds

(a) $((B_{\mathcal{M}} - F) \cup (F - B_{\mathcal{M}})) \cap AT^+ \subsetneq ((B_{\mathcal{M}} - F') \cup (F' - B_{\mathcal{M}})) \cap AT^+$

(b) $((B_{\mathcal{M}} - F) \cup (F - B_{\mathcal{M}})) \cap AT^+ = ((B_{\mathcal{M}} - F') \cup (F' - B_{\mathcal{M}})) \cap AT^+$ and $((B_{\mathcal{M}} - F) \cup (F - B_{\mathcal{M}})) \cap (AT_X^+ - AT^+) \subsetneq ((B_{\mathcal{M}} - F') \cup (F' - B_{\mathcal{M}})) \cap (AT_X^+ - AT^+)$

Apply this revised semantics to Kennedy's example. Assume that the period in question is $[0, t_0]$ where the two assassins received the command from boss at 0 and then get ready. The conspiracy is simplified as the command received by the Oswald and Aswald since there is a correlation between the former and the later, that is, the later is a necessary consequence of the former and the later wouldn't have happened if without the former. Moreover, at time t_1, Kennedy starts to move (without loss of generality, assume the movement is linear and at speed 1). When Kennedy moves to distance d_1, it comes into Oswald's range of fire and Oswald fires immediately. If Kennedy survives Oswald's short, Aswald would fire and kill him when he moves to distance d_2 such that $t_1 + d_2 \in (t_1 + d_1, t_0]$.

The domains are as follows

$$
\begin{aligned}
D_e = {} & \{receive, Ofire, Afire, start\} \\
D_f = {} & \{Oready, Aready, alive, dead, moving, distance(x)\} \\
Dom = {} & \{Inl(alive), Inl(dead), Inl(Oready), Inl(Aready), Inl(moving), Hap(receive, t), \\
& Hap(Ofire, t), Hap(Afire, t), Hap(start, t), Ini(receive, Oready, t), \\
& Ini(receive, Aready, t), Ter(Ofire, alive, t), Ini(Ofire, dead, t), Ter(Afire, alive, t), \\
& Ini(Afire, dead, t), Ini(start, moving, t), Inl(distance(x)), Tra(moving, t, distance(x), d), \\
& Rel(start, distance(), t), Cli(t, alive, t'), Cli(t, distance(x), t'), Cli(t, Oready, t') \\
& Cli(t, Aready, t'), Cli(t, moving, t'), Cli(t, dead, t')\}
\end{aligned}
$$

where *receive* is short for *The assassins receive the command* and the others are such that

$Ofire$	for	*Oswald fires*
$Afire$	for	*Aswald fires*
$start$	for	*Kennedy starts to move*
$Oready$	for	*Oswald is ready for shooting*
$Aready$	for	*Aswald is ready for shooting*
$alive$	for	*Kennedy is alive*
$dead$	for	*Kennedy is dead*
$moving$	for	*Kennedy is moving linearly at speed 1*
$distance(x)$	for	*The distance of Kennedy's movement is* x

L_C contains the following causal laws

- $\neg Hol(Oready, t) \rightarrow Ini(receive, Oready, t)$
- $\neg Hol(Aready, t) \rightarrow Ini(receive, Oready, t)$
- $\neg Hol(moving, t) \rightarrow Ini(start, moving, t)$
- $\neg Hol(moving, t) \rightarrow Rel(start, distance(), t)$
- $Hol(distance(x), t) \rightarrow Tra(moving, t, distance(x + d), d)$
- $Hol(alive, t) \wedge Hol(X, t) \rightarrow Ter(Ofire, alive, t)$
- $Hol(alive, t) \wedge Hol(X, t) \rightarrow Ini(Ofire, dead, t)$
- $Hol(alive, t) \wedge Hol(Oready, t) \wedge Hol(distance(d_1), t) \rightarrow Hap(Ofires, t)$
- $Hol(alive, t) \rightarrow Ter(Afire, alive, t)$
- $Hol(alive, t) \rightarrow Ini(Afire, dead, t)$
- $Hol(alive, t) \wedge Hol(Aready, t) \wedge Hol(distance(d_2), t) \rightarrow Hap(Afires, t)$

The completion can be spelled out in a routine way and thus omitted here. X holds at time $t_1 + d_1$, which grounds the success of Oswald's shooting. But no information is available about the truth value of $Hol(X, t)$ for all $t \in [0, t_1 + d_1) \cup (t_1 + d_1, t_0]$, so there is no unique model representing the story. Pick an arbitrary $\mathcal{M}$ of them, the basis $B_{\mathcal{M}}$ consists of the following atomic facts by definition.

$\{\neg Inl(Oready), \neg Inl(Aready), Inl(alive), \neg Inl(dead), \neg Inl(moving), Inl(distance(0))$
$Hap(receive, 0), Hap(start, t_1), Hol(X, t_1 + d_1)\} \cup$
$\{\neg Hap(receive, t) \mid t \in (0, t_1]\} \cup \{\neg Hap(start, t) \mid t \in [0, t_1) \cup (t_1, t_0]\}$

and those $(\neg)Hol(X, t)$ for $t \in [0, t_1 + d_1) \cup (t_1 + d_1, t_0]$. The integrity constraint apart from (1a), (1b), (2a) and (2b) is

$$Hol(distance(x), t) \wedge Hol(distance(y), t) \rightarrow x = y$$

Various modifications of $B_{\mathcal{M}}$ are sufficient for the counterfactual assumption *if Oswald hadn't killed Kennedy* (formally, $\neg(Hap(Ofire, t) \wedge Ini(Ofire, dead, t))$ for all $t \in [0, t_0]$), e.g. Kennedy doesn't start to move or the assassins haven't received the command. Any modification involving the change of atomic facts other than those of the form $(\neg)Hol(X, t)$ will be less similar to $\mathcal{M}$, and mere modifications of $(\neg)Hol(X, t)$ will necessarily concerns $(\neg)Hol(X, t_1 + d_1)$. Actually, if those independent particular facts not of the form $(\neg)Hol(X, t)$ are fixed, only $(\neg)Hol(X, t_1 + d_1)$ can influence the success or failure of Oswald because Kennedy reaches distance d_1 precisely at time $t_1 + d_1$. It follows that the unique minimal modification of $B_{\mathcal{M}}$ is abandoning $Hol(X, t_1 + d_1)$ and adding its negation $\neg Hol(X, t_1 + d_1)$. Let

$$B' = B_{\mathcal{M}} - \{Hol(X, t_1 + d_1)\} \cup \{\neg Hol(X, t_1 + d_1)\}$$

it could be checked routinely that

$$B' \cup L_D \cup IC \models Hap(Afire, t_1 + d_2) \wedge Ini(Afire, dead, t_1 + d_2)$$

that is, Aswald would have killed Kennedy at time $t_1 + d_2$.

6 Conclusion

Though complicated, causation can be specified on the basis of the acquisition of causal laws and particular facts by making use of information about timing. Event calculus serves as a pretty fine-grained framework for this task. Given the mechanism of identification of causation, the semantics of counterfactuals has been formulated under event calculus. With sufficient details about the causal structure underlying each situation, the semantics can be used to determine the truth value of counterfactuals precisely, e.g. those about neuron networks; while with merely partial information, the semantics can also preserve the vagueness which does exist in practical counterfactual reasoning, e.g. those about people's daily activities.

On the contrary, it has also been shown that there is probably a gap between the poor differentiating power of counterfactuals and the complication of causal structures, which permanently threaten the soundness of counterfactual analysis of causation.

Given the contrast between the two approaches, it can be concluded that it's causation that underlies the semantics of causation rather than the converse.

References

[1] J. Bennett. Event causation: the counterfactual analysis. *Philosophical Perspectives*, 1:367–386, 1987.

[2] D. D. Cummins. Naive theories and causal deduction. *Memory & Cognition*, 23(5):646–658, 1995.

[3] N. Hall. Two concepts of causation. In *Causation and Counterfactuals*. the MIT Press, 2004.

[4] N. Hall J. Collins and L. A. Paul. Counterfactuals and causation: history, problems, and prospects. In *Causation and Counterfactuals*. The MIT Press, 2004.

[5] A. Kratzer. An investigation of the lumps of thought. *Linguistics and Philosophy*, 12:607–653, 1989.

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

[7] D. Lewis. *Counterfactuals*. Basil Blackwell, Oxford, 1973.

[8] D. Lewis. Counterfactual dependence and time's arrow. *Noûs*, 13 (4):455–476, 1979.

[9] D. Lewis. Causation as influence. In *Causation and Counterfactuals*. The MIT Press, 2004.

[10] L. A. Paul. Keeping track of the time: emending the counterfactuals analysis of causation. *Analysis*, 58 (3):191–198, 1998.

[11] K. Schulz. *Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals*. PhD thesis, University of Amsterdam, 2007.

[12] K. Schulz. "if you'd wiggled a, then b would've changed": Causality and counterfactual conditionals. *Synthese*, 179 (2):239–251, 2010.

[13] R. Stalnaker. A theory of conditionals. *Studies in Logical Theory*, pages 98–112, 1968.

[14] A. Neeleman & H. van de Koot. The linguistic expression of causation. In *UCLWPL*. 2010.

[15] M. van Lambalgen & F. Hamm. *The Proper Treatment of Events*. Blackwell, 2004.

[16] K. Stenning & M. van Lambalgen. *Human Reasoning and Cognitive Science*. The MIT Press, 2008.

[17] F. Veltman. Making counterfactual assumptions. *Journal of Semantics*, 22 (2):159–180, 2005.