

Configurations for Inference between Causal Statements

Philippe Besnard¹, Marie-Odile Cordier², and Yves Moinard³

¹ CNRS, IRIT, Université Paul Sabatier
118 route de Narbonne, 31062 Toulouse cedex, France
`besnard@irit.fr`

² Université de Rennes I, IRISA
Campus de Beaulieu, 35042 Rennes cedex, France
`cordier@irisa.fr`

³ INRIA, IRISA
Campus de Beaulieu, 35042 Rennes cedex, France
`moinard@irisa.fr`

Abstract. When dealing with a cause, cases involving some effect due to that cause are precious as such cases contribute to what the cause is. They must be reasoned upon if inference about causes is to take place. It thus seems like a good logic for causes would arise from a semantics based on collections of cases, to be called configurations, that gather instances of a given cause yielding some effect(s). Two crucial features of this analysis of causation are transitivity, which is endorsed here, and the event-based formulation, which is given up here in favor of a fact-based approach. A reason is that the logic proposed is ultimately meant to deal with both deduction (given a cause, what is to hold?) and abduction (given the facts, what could be the cause?) thus paving the way to the inference of explanations. The logic developed is shown to enjoy many desirable traits. These traits form a basic kernel which can be modified but which cannot be extended significantly without losing the adequacy with the nature of causation rules.

1 Motivation

Causation as entertained here concerns the usual relationship that may hold between states of affairs (thus departing from the concept favored by many people who focus on events —up to the point of insisting that causation is based on events⁴). For the relationship to hold, the cause must be what brings about the effect. Moreover, the account here is that the cause always brings about the effect (ruling out that “smoking causes cancer” may count as true). Such a strict reading makes a probabilistic interpretation of causation [16] (whether the probabilities are subjective or not) less appealing but is more sympathetic to

⁴ The event-based approach is far from being uncontroversial: For instance, one usually says that the gravitational attraction of the moon (and of the sun) causes tides even if no event, but a principle, is indicated as the cause.

the so-called counterfactual analysis of causation [11]. As is well-known, the idea of causal dependence is thus originally stated in terms of events (where c and e are two distinct possible events, e causally depends on c if and only if c occurs counterfactually implies e occurs and c does not occur counterfactually implies e does not occur, or if it does occur, it is due to some other rule), but facts have been reinstated [14]. Actually, statements might accommodate the technicalities pinpointed in [14] that insists on dealing with facts in a general sense (in order to capture causation involving negative existential facts for example: “she didn’t get a speeding ticket because she was able to slow down early enough”).

Although the counterfactual analysis of causation is meant to address the issue of truth for causation, the aim here is *not* a logic of causation: The purpose is not to give truth conditions for causal statements in terms that do not themselves appeal to causal concepts. The aim is only to provide truth-preserving conditions between causal statements. In particular, the logic will distinguish between a pair of statements essentially related as cause and effect and a pair of statements which are merely effects of a common cause (correlations being not confused with instances of causation), something conceptually troublesome for probabilistic approaches.

The logic will allow for transitivity as far as deterministic causes are concerned. No discussion on such a controversial topic is included. Instead, a single comment is as follows: Prominent authors in the defence of transitivity for causation include Lewis [13] and Hall [8] (see also Yablo [19]).

2 Introduction

When dealing with a cause (e.g., looking for something which could explain certain facts), cases involving some effect due to that cause are precious as such cases contribute to what the cause is. Accordingly, they must be reasoned upon if inference about causes is to take place. It thus seems like a good logic for causes would arise from a semantics based on collections of cases, to be called configurations that gather instances of a given cause yielding some effect(s).

The setting of configurations is what permits to discriminate correlations from instances of causes: If α and β are equivalent for example, then every situation admits both or none; when it comes to describing causation with respect to α for example, nothing requires to mention β for the reason that configurations are only supposed to take causation but not truth into account.

In the formalism to be introduced below, δ *causes* $\langle\beta\rangle$ means that δ causes the single effect β . Should α and β be equivalent, every situation is such that both α and β are true or none is; however, a configuration for δ must mention β as δ causes β but it need not be so for α (because δ does not cause α , regardless of the fact that α is true exactly when β is).

More generally, α *causes* $\langle\beta_1, \dots, \beta_n\rangle$ means that the effects caused by α consist exactly of β_1, \dots, β_n where each β_i is *the* effect in the set $\{\beta_1, \dots, \beta_n\}$, caused by a certain occurrence of α . Importantly, it is thus possible to express the possibility for a cause to have alternative effects like in the following example:

“Turning the wheel causes the car to go left or right”. The outcome is that amongst the configurations for “turning the wheel”, some include “the car goes to left” and some include “the car goes to right”⁵.

We propose a framework which is minimal in that only a few properties (all of which seemingly uncontroversial) are imposed upon it. Additional technicalities may later enrich the framework, depending on what application domain is considered. The *causes* operator is meant to be used with respect to an intended domain. A formula α *causes* $\langle \beta_1, \dots, \beta_n \rangle$ is supposed to mean that part of our causal knowledge is that α has $\{\beta_1, \dots, \beta_n\}$ as an exhaustive set of alternative effects. Here, “exhaustive” means that $\{\beta_1, \dots, \beta_n\}$ accurately describes a set of alternative effects of a given kind (not precluding the existence of other kinds of effects, to be specified by means of other causal formulas).

Drastically discretizing temperatures in order to keep things simple, the example *flu causes* $\langle t_{38}, t_{39}, \dots, t_{41} \rangle$ illustrates what is meant here: Adding t_{37} , or removing t_{38} would modify the intended meaning for the disease called *flu*. However, adding formulas such as *flu causes* $\langle fatigue \rangle$ is possible.

The general idea is that, from such causal formulas together with classical formulas, some consequences are to be formally derived according to the contents of the causal configurations.

According to our principled assumption requiring that causes-effects relations are captured through configurations, the following properties should not hold:

- δ *causes* $\langle \alpha \rangle \rightarrow \delta$ *causes* $\langle \alpha, \beta \rangle$ should be untrue in general – and of course so does the converse formula δ *causes* $\langle \alpha, \beta \rangle \rightarrow \delta$ *causes* $\langle \beta \rangle$.
- Neither causation nor effect should be strongly related to classical implication: $\beta \rightarrow \delta$ should entail neither δ *causes* $\langle \alpha \rangle \rightarrow \beta$ *causes* $\langle \alpha \rangle$ nor α *causes* $\langle \beta \rangle \rightarrow \alpha$ *causes* $\langle \delta \rangle$.
- Generally, δ *causes* $\langle \delta \rangle$ (reflexivity) should fail, even if it should be possible to make it hold when necessary, in special cases involving a cycling causal phenomenon.
- Chaining of nondeterministic causes is undesirable: δ *causes* $\langle \alpha, \beta \rangle$ together with β *causes* $\langle \gamma, \epsilon \rangle$ need not entail δ *causes* $\langle \alpha, \gamma, \epsilon \rangle$. The idea here is that β in full generality may cause either γ or ϵ but in the particular context where β occurs due to δ , it might well be that only γ (for instance) can happen. However, the idea of transitivity should remain, the precise formulation being postponed to the technical presentation given below in section 5. In particular, chaining of deterministic causes is desired in the form of the following property: δ *causes* $\langle \alpha, \gamma \rangle$ and α *causes* $\langle \beta \rangle$ should infer δ *causes* $\langle \beta, \gamma \rangle$.

⁵ A question is whether a cause with alternative effects should be formalized as a cause with a single disjunctive effect. This is not the solution envisioned here because the notion of a single disjunctive effect seems somewhat shaky (it is assumed that an effect is described by means of a statement). About the well-known example that Suzy may throw a rock at a glass bottle and Billy may throw a rock at another glass bottle, some authors (Collins, Hall and Paul [5]) deny that the disjunctive item “Suzy’s bottle shatters or Billy’s bottle shatters” is apt to be caused: There is no such thing as a disjunctive effect.

3 Formal Definitions

The language of classical propositional logic is extended with formulas having multiple arguments: α *causes* $\langle\beta\rangle$ means that α causes the single effect β , and α *causes* $\langle\beta_1, \dots, \beta_n\rangle$ (where n is finite) means that one of these β_i is a possible effect caused by a certain occurrence of α . In order to keep causal statements simple, α and β_1, \dots, β_n are atomic formulas of classical propositional logic.

Causal formulas are defined as follows, where α, β_i are propositional symbols:

1. Each propositional symbol (*propositional atom*) is a causal formula,
2. Each *causal atom* α *causes* $\langle\beta_1, \dots, \beta_n\rangle$ is a causal formula.
3. If φ_1 and φ_2 are causal formulas, so are $\neg\varphi_1$, $\varphi_1 \wedge \varphi_2$, $\varphi_1 \vee \varphi_2$, $\varphi_1 \rightarrow \varphi_2$ and $\varphi_1 \leftrightarrow \varphi_2$.

A *propositional formula* is a causal formula without any causal atom, and “*formula*” will often be used instead of “*causal formula*”. A [*causal*] *theory* CT is a set of formulas, the set of the propositional formulas in CT being denoted by W . As an illustration, CT may consist of $\neg(\alpha$ *causes* $\langle\beta\rangle)$ and $(\alpha$ *causes* $\langle\gamma\rangle) \rightarrow (\beta$ *causes* $\langle\gamma, \delta\rangle)$ together with $\neg(\beta \wedge \gamma \wedge \delta)$ (which makes W) where $\alpha, \beta, \gamma, \delta$ are propositional atoms.

The notion of *configuration* is to be used to specify the cases of reference between a cause and its effects. Letting \mathcal{I} denote the set of interpretations in classical propositional logic, a *configuration* is a set of principal filters from $2^{\mathcal{I}}$ (hence an element of a configuration is the set of all the subsets of \mathcal{I} which contain some given subset of \mathcal{I}). Since a set of interpretations is routinely identified with any (propositional) formula satisfied in exactly that set, a *configuration* can be assimilated with a set of conjunctions of propositional atoms.

Notice that the conjunction of 0 atoms, that is the true formula \top (which corresponds to the full set $2^{\mathcal{I}}$, which is a principal filter), is an eligible element of a configuration, while the false formula \perp (which corresponds to the empty set, which is not a filter), is not allowed here. Here is a simple example:

Flu causes some high temperature (either 38° or 39° or 40°) and 40° causes shiver: $\varphi_1 = \text{flu causes } \langle t38, t39, t40 \rangle$, $\varphi_2 = t40$ *causes* $\langle \text{shiver} \rangle$.
Here are three examples of configurations:
 $S_1 = \{t38, t39, t40 \wedge \text{shiver}\}$, $S_2 = \{\text{shiver}\}$, $S_3 = \{\top\}$.

Satisfaction with respect to an interpretation of classical propositional logic is denoted by means of the symbol \models (e.g., $I \models \alpha$). which is also used to denote the relation of logical consequence from classical propositional logic.

A *causal interpretation* is a pair $\langle \mathcal{S}, I \rangle$ where I is an interpretation in classical propositional logic, \mathcal{S} is a family (indexed by the propositional atoms) of configurations. In symbols, $I \in \mathcal{I}$ and $\mathcal{S} = \{S_\alpha, S_\beta, \dots\}$ where each $S_\alpha \subseteq 2^{\mathcal{I}}$ is a set of principal filters and α, β, \dots is a list of all propositional atoms.

Definition 1. *Causal satisfaction relation and causal inference*

A causal interpretation $\mathcal{C} = \langle \mathcal{S}, I \rangle$ satisfies a formula γ (written $\mathcal{C} \Vdash \gamma$) according to the following recursive rules:

$$\mathcal{C} \Vdash \neg\delta \quad \text{if} \quad \mathcal{C} \not\Vdash \delta \quad (1)$$

$$\mathcal{C} \Vdash \delta \vee \epsilon \quad \text{if} \quad \mathcal{C} \Vdash \delta \text{ or } \mathcal{C} \Vdash \epsilon \quad (2)$$

$$\mathcal{C} \Vdash \delta \rightarrow \epsilon \quad \text{if} \quad \mathcal{C} \not\Vdash \delta \text{ or } \mathcal{C} \Vdash \epsilon \quad (3)$$

$$\mathcal{C} \Vdash \delta \wedge \epsilon \quad \text{if} \quad \mathcal{C} \Vdash \delta \text{ and } \mathcal{C} \Vdash \epsilon \quad (4)$$

$$\mathcal{C} \Vdash \alpha \quad \text{if} \quad I \models \alpha \text{ for } \alpha \text{ propositional atom} \quad (5)$$

$$\mathcal{C} \Vdash \alpha \text{ causes } \langle \beta_1, \dots, \beta_n \rangle \quad \text{if} \quad \begin{cases} I \not\models \alpha \wedge \neg\beta_1 \wedge \dots \wedge \neg\beta_n & \text{and} \\ \forall X \in S_\alpha \exists \beta_i \exists Y \in S_{\beta_i} X \models \beta_i \wedge Y, & \\ \forall \beta_i \exists X \in S_\alpha \exists Y \in S_{\beta_i} X \models \beta_i \wedge Y. & \end{cases} \quad (6)$$

We define *causal inference*, also denoted \Vdash , from the causal satisfaction relation as usual: $CT \Vdash \gamma$ holds iff all models of CT are models of $\{\gamma\}$.

As the configuration S_δ lists the cases describing the effects of a cause δ , the second condition $\forall X \in S_\alpha \exists \beta_i \dots$ in (6) expresses that there is no case in which α causes none of β_1, \dots, β_n . The third condition $\forall \beta_i \exists X \in S_\alpha \dots$ expresses conversely that each of β_1, \dots, β_n does exist as an effect of α (this conditions reduces to $S_\alpha \neq \emptyset$ in case of a single effect $\langle \beta_1 \rangle$). For the reader to better grasp the intuitions underlying the above definition, let us continue our “flu” example:

Take $\mathcal{C} = \langle \mathcal{S}, I \rangle$ where $I = \emptyset$ (i.e., $I \models \neg flu$ and so on) and \mathcal{S} such that $S_{flu} = \{t38, t39, t40 \wedge shiver\}$, $S_{t40} = \{shiver\}$, $S_{t38} = S_{t39} = S_{shiver} = \{\top\}$. Then, \mathcal{C} is a model of $\{\varphi_1, \varphi_2\}$ (notice the mandatory occurrence of *shiver* together with *t40* in S_{flu}). Let us verify $\mathcal{C} \Vdash t40 \text{ causes } \langle shiver \rangle$:

- As to the first condition, $I \not\models t40 \wedge \neg shiver$ because $I \models \neg t40$.
- The second condition $\forall X \in S_{t40} \exists Y \in S_{shiver} X \models shiver \wedge Y$ is then instantiated by $X = shiver$ and $Y = \top$.
- The third condition reduces here to $S_{t40} \neq \emptyset$ (single effect).

Let us check $\mathcal{C} \not\Vdash t40 \text{ causes } \langle t40 \rangle$: The second condition fails: X must be *shiver* here and, since $shiver \not\models t40$, we cannot get $X \models t40 \wedge Y$.

The semantics just presented bears some similarity with semantics involving a selection function for conditionals (among others, a version in [6] is: models are equipped with a family of functions indexed by \mathcal{I} from the set of formulas to the powerset of \mathcal{I} i.e. something fairly close to configurations). It does not come as a surprise that specifying causation-based cases shares some technical aspects with specifying counterfactual cases (would-be states of affairs). Of course, there cannot be an algebraic semantics in the usual sense that a Boolean algebra is endowed with an extra binary operation. As with logics failing substitution principles, some technical tricks would have to be used instead as in [15].

4 A few Features of this Semantics

4.1 Two small typical examples

Let us consider the following situation: $\alpha \text{ causes } \langle \beta \rangle$, $\alpha \text{ causes } \langle \gamma \rangle$. (S1)

We are looking for a model $\mathcal{C} = \langle \mathcal{S}, I \rangle$ of (S1). As for I , all we need is a model of the two formulas $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$. Let us choose $I = \{\alpha, \beta, \gamma\}$ (all propositional atoms true). As for \mathcal{S} , we need β and γ in each element of S_α . Here is a possibility: $S_\alpha = \{\beta \wedge \gamma\}$, $S_\beta = S_\gamma = \{\top\}$.

This model satisfies also the formula α causes $\langle \beta, \gamma \rangle$, and in fact each model of (S1) satisfies α causes $\langle \beta, \gamma \rangle$, meaning that we have: (S1) $\Vdash \alpha$ causes $\langle \beta, \gamma \rangle$.

Here is another typical situation: α causes $\langle \beta \rangle$, β causes $\langle \gamma \rangle$. (S2)

The model given above for (S1) falsifies β causes $\langle \gamma \rangle$. Indeed, γ must be in each element of S_β . Then, $\beta \wedge \gamma$ must be in each element of S_α . We can choose $I = \{\alpha, \beta, \gamma\}$ again, together with: $S_\alpha = \{\beta \wedge \gamma\}$, $S_\beta = \{\gamma\}$, $S_\gamma = \{\top\}$. Each model of (S2) satisfies α causes $\langle \gamma \rangle$, meaning that we get (S2) $\Vdash \alpha$ causes $\langle \gamma \rangle$. Remind that we consider transitivity as a desirable feature.

4.2 Where S_α is the Empty Set

If $S_\alpha = \emptyset$ in a causal interpretation $\mathcal{C} = \langle \mathcal{S}, I \rangle$, then $\mathcal{C} \Vdash \neg(\alpha$ causes $\langle \beta_0 \cdots, \beta_n \rangle)$ and $\mathcal{C} \Vdash \neg(\beta_0$ causes $\langle \alpha, \beta_1, \cdots, \beta_n \rangle)$: α is neither a “cause” nor an “effect”.

4.3 Irreflexivity

The condition for $\mathcal{C} \Vdash \delta$ causes $\langle \delta \rangle$ simplifies as $S_\delta \neq \emptyset$ and $\forall X \in S_\delta \ X \models \delta$.

This is why the above semantics invalidates δ causes $\langle \delta \rangle$.

Moreover, δ causes $\langle \gamma \rangle$ entails neither δ causes $\langle \delta \rangle$ nor γ causes $\langle \gamma \rangle$.

4.4 Transitivity

Chains of deterministic causes are admitted as shown by the valid inference

$$\text{From } \delta \text{ causes } \langle \alpha \rangle \text{ and } \alpha \text{ causes } \langle \beta \rangle \text{ infer } \delta \text{ causes } \langle \beta \rangle.$$

We have already stated this result with (S2) in §4.1. When it comes to chains of nondeterministic causes, the pertinent result is postponed to section 5 below. In particular, as expected (cf Introduction) the following inference is invalid:

$$\text{From } \delta \text{ causes } \langle \alpha, \beta \rangle \text{ and } \beta \text{ causes } \langle \gamma, \epsilon \rangle \text{ infer } \delta \text{ causes } \langle \alpha, \gamma, \epsilon \rangle.$$

4.5 Sets of Effects Need Not Be Minimal in Causal Atoms

From δ causes $\langle \alpha_1, \cdots, \alpha_m \rangle$ and δ causes $\langle \beta_1, \cdots, \beta_n \rangle$,
we can infer δ causes $\langle \alpha_1, \cdots, \alpha_m, \beta_{i_1}, \cdots, \beta_{i_k} \rangle$
for any list $\beta_{i_1}, \cdots, \beta_{i_k}$ of elements of the set $\{\beta_1, \cdots, \beta_n\}$.

This property, which generalizes situation (S1) in §4.1, shows that cumulative effects are turned into disjunctive effects, so to speak, which is in accordance with the classical “and implies or”, here applied to effects. This feature of the semantics shows that *causal atoms* are not absolutely *atomic*, but this was already clear from their definition, which involves “atoms” of different size.

Remind that having δ causes $\langle \alpha_1, \dots, \alpha_m \rangle$ and δ causes $\langle \beta_1, \dots, \beta_n \rangle$ is not exceptional. This is exemplified by subsection 4.4 where we get: if δ causes $\langle \alpha \rangle$ and α causes $\langle \beta \rangle$, then δ causes $\langle \beta \rangle$ (thus δ causes $\langle \alpha, \beta \rangle$).

4.6 Contradictory Effect

Our formalism does not allow causal atoms involving directly the false formula. An empty list $\langle \rangle$ can be assimilated to \perp but, since we have excluded \perp as an eligible element of a configuration, the second condition of (6) in Definition 1 cannot be satisfied: $\Vdash \neg(\delta \text{ causes } \langle \rangle)$. Thus, we forbid empty lists $\langle \rangle$, leaving the introduction of the (single) contradictory effect for future work. This would in particular simulate some “causal negation”, thus extending significantly the expressive power, but it is not a trivial matter.

4.7 Links with Logical Consequence

Deduction theorem: Due to the (rather traditional and classical) definition of the semantics and of the inference relation, the deduction theorem holds:

$$CT \cup \{\varphi\} \Vdash \psi \quad \text{iff} \quad CT \Vdash \varphi \rightarrow \psi.$$

Remark 1. Since condition $I \not\models \delta \wedge \neg\gamma_1 \wedge \dots \wedge \neg\gamma_n$ is equivalent to $I \models \delta \rightarrow (\gamma_1 \vee \dots \vee \gamma_n)$, the semantics validates the following inference:

$$\delta \text{ causes } \langle \beta_1, \dots, \beta_n \rangle \Vdash \delta \rightarrow (\beta_1 \vee \dots \vee \beta_n)$$

which, by the deduction theorem, is equivalent to

$$\Vdash (\delta \text{ causes } \langle \beta_1, \dots, \beta_n \rangle) \rightarrow (\delta \rightarrow (\beta_1 \vee \dots \vee \beta_n)).$$

Remark 2. Logical consequence fails in general to carry over to effects. Actually, that $\alpha \rightarrow \beta$ is a consequence of a theory CT does not entail that $\delta \text{ causes } \langle \alpha \rangle \rightarrow \delta \text{ causes } \langle \beta \rangle$ is a consequence of CT . Technically, the reason is that $\delta \text{ causes } \langle \alpha \rangle$ imposes no condition on any configuration about β .

Remark 3. From Remark 1 we get that if δ causes α then whatever entails δ also entails α , but it need not cause α . It seems right that causation not be strongly related to logical consequence. Here is an illustration: It is certainly true that “being a compulsive gambler causes me to lose lots of money” but it seems more controversial to hold that “being a compulsive gambler and feeling sleepy causes me to lose lots of money”. The above semantics fails

$$\text{if } \alpha \text{ causes } \langle \gamma \rangle \text{ and } \delta \rightarrow \alpha \text{ then } \delta \text{ causes } \langle \gamma \rangle,$$

which in turn invalidates

$$\text{if } \alpha \text{ causes } \langle \gamma \rangle, \beta \text{ causes } \langle \gamma \rangle, \text{ and } \delta \rightarrow \alpha \vee \beta, \text{ then } \delta \text{ causes } \langle \gamma \rangle. \quad (7)$$

Remark 4. A related invalid principle is

if δ causes $\langle \alpha \rangle$ and $\alpha \leftrightarrow \beta$ then δ causes $\langle \beta \rangle$.

This principle becomes valid under the following constraint

$$\text{If } W \models \alpha \leftrightarrow \beta \text{ then } \left(\begin{array}{l} \forall X \in S_\delta \exists Y \in S_\alpha X \models Y \wedge \alpha \Rightarrow \\ \forall X \in S_\delta \exists Z \in S_\beta X \models Z \wedge \beta \end{array} \right) \quad (\text{EE})$$

Remark 5. Similarly, the above semantics invalidates

If δ causes $\langle \alpha \rangle$ and $\delta \leftrightarrow \eta$ then η causes $\langle \alpha \rangle$

on the intuitive grounds that a cause is (roughly speaking) a reason for some effect(s) to happen whereas being true simultaneously with the cause is not enough for also being a reason for the effect(s).

4.8 Causes from New Premises Are Impossible

We can never infer a causal atom α causes $\langle \beta_1, \dots, \beta_n \rangle$ from a theory CT which does not contain already, directly or indirectly, some causal atom α causes $\langle \gamma_1, \dots, \gamma_m \rangle$. By “indirectly” here we mean allowing only “classical boolean inference”, where causal atoms are dealt with as if they were new propositional atoms, e.g. inferring α causes $\langle \epsilon \rangle$ from δ causes $\langle \epsilon \rangle$ and δ causes $\langle \epsilon \rangle \rightarrow \alpha$ causes $\langle \epsilon \rangle$.

5 Proof System

The proof system \vdash_c consists of any proof system for classical propositional logic extended with the following schemata

1. δ causes $\langle \gamma_1, \gamma_1, \gamma_2, \dots, \gamma_n \rangle \leftrightarrow \delta$ causes $\langle \gamma_1, \gamma_2, \dots, \gamma_n \rangle$.
2. δ causes $\langle \gamma_1, \dots, \gamma_{i-1}, \gamma_i, \dots, \gamma_n \rangle \rightarrow \delta$ causes $\langle \gamma_1, \dots, \gamma_i, \gamma_{i-1}, \dots, \gamma_n \rangle$.
3. δ causes $\langle \gamma_1, \dots, \gamma_n \rangle \rightarrow (\delta \rightarrow \gamma_1 \vee \dots \vee \gamma_n)$.
4. δ causes $\langle \gamma_1, \dots, \gamma_n \rangle \wedge \gamma_1$ causes $\langle \alpha_1, \dots, \alpha_m \rangle \rightarrow$
 $\bigvee_R \delta$ causes $\langle \alpha_{i_1}, \dots, \alpha_{i_k}, \gamma_2, \dots, \gamma_n \rangle$ where the range R is
 $\emptyset \neq \{\alpha_{i_1}, \dots, \alpha_{i_k}\} \subseteq \{\alpha_1, \dots, \alpha_m\}$.
5. δ causes $\langle \gamma_1, \dots, \gamma_n \rangle \wedge \delta$ causes $\langle \alpha_1, \dots, \alpha_m \rangle \rightarrow$
 δ causes $\langle \gamma_1, \dots, \gamma_n, \alpha_{i_1}, \dots, \alpha_{i_k} \rangle$ where each α_{i_j} is in $\{\alpha_1, \dots, \alpha_m\}$.

Schemas 1 and 2 just say that the lists $\langle \gamma_1, \gamma_1, \gamma_2, \dots, \gamma_n \rangle$ must in fact be considered as sets of formulas. Schema 3 refers to the result of *Remark 1* in § 4.7. Schema 4 describes what remains of transitivity (cf § 4.4). Schema 5 ensures that we get the result mentioned in § 4.5.

It is easy to prove that the logic presented in this text is sound, while completeness remains a conjecture:

Theorem 2. *If $CT \vdash_c \varphi$ then $CT \Vdash \varphi$.*

Two elementary typical examples of using this proof system are provided by the two situations of §4.1:

Case (S1): Point 5 gives $(\alpha \text{ causes } \langle \beta \rangle \wedge \alpha \text{ causes } \langle \gamma \rangle) \rightarrow \alpha \text{ causes } \langle \beta, \gamma \rangle$.

Case (S2): Point 4 gives $(\alpha \text{ causes } \langle \beta \rangle \wedge \beta \text{ causes } \langle \gamma \rangle) \rightarrow \alpha \text{ causes } \langle \gamma \rangle$.

6 Comments about Transitivity

6.1 A few Valid and Invalid Principles

Here are two typical instances of schema 4:

$$\delta \text{ causes } \langle \gamma_1, \dots, \gamma_n \rangle \wedge \gamma_1 \text{ causes } \langle \alpha \rangle \rightarrow \delta \text{ causes } \langle \alpha, \gamma_2, \dots, \gamma_n \rangle;$$

$$\delta \text{ causes } \langle \gamma \rangle \wedge \gamma \text{ causes } \langle \alpha, \beta \rangle \rightarrow \begin{pmatrix} \delta \text{ causes } \langle \alpha \rangle & \vee \\ \delta \text{ causes } \langle \beta \rangle & \vee \\ \delta \text{ causes } \langle \alpha, \beta \rangle & \end{pmatrix}$$

The following are three consequences of these results, which concern what could be called “causal equivalence”. We suppose that a theory contains the formula

$$(\alpha \text{ causes } \langle \beta \rangle) \wedge (\beta \text{ causes } \langle \alpha \rangle). \quad (8)$$

Then, we get:

$$\text{If } \delta \text{ causes } \langle \alpha \rangle, \text{ then } \delta \text{ causes } \langle \beta \rangle; \quad (9)$$

$$\text{If } \delta \text{ causes } \langle \alpha, \alpha_1, \dots, \alpha_n \rangle, \text{ then } \delta \text{ causes } \langle \beta, \alpha_1, \dots, \alpha_n \rangle; \quad (10)$$

$$\text{If } \alpha \text{ causes } \langle \gamma \rangle, \text{ then } \beta \text{ causes } \langle \gamma \rangle. \quad (11)$$

These results, which must be compared with Remark 4 in § 4.7 for cases (9) and (10), and with Remark 5 for case (11), show that “causal equivalence” is stronger than boolean equivalence, as expected. Notice however that, still supposing formula (8), the following principle remains invalid:

$$\text{If } \alpha \text{ causes } \langle \gamma_1, \dots, \gamma_n \rangle, \text{ then } \beta \text{ causes } \langle \gamma_1, \dots, \gamma_n \rangle.$$

We get no more than the disjunction obtained by point (4) in § 5:

If $\alpha \text{ causes } \langle \gamma_1, \dots, \gamma_n \rangle$, then $\bigvee_{\emptyset \neq \{\gamma_{i_1}, \dots, \gamma_{i_k}\} \subseteq \{\gamma_1, \dots, \gamma_n\}} (\beta \text{ causes } \langle \gamma_{i_1}, \dots, \gamma_{i_k} \rangle)$.

Here, “causal equivalence” is not stronger than the causal atom $\beta \text{ causes } \langle \alpha \rangle$.

6.2 Enlarging the Semantics

Let us suppose that we have the following causal theory CT_1 :

$$\alpha \text{ causes } \langle \beta \rangle, \beta \text{ causes } \langle \gamma, \epsilon \rangle, \alpha \text{ causes } \langle \epsilon' \rangle; \quad \epsilon' \rightarrow \neg \epsilon.$$

Here are a few causal consequences of CT_1 :

$$\alpha \text{ causes } \langle \beta, \epsilon' \rangle, \quad \alpha \text{ causes } \langle \gamma \rangle \vee \alpha \text{ causes } \langle \epsilon \rangle \vee \alpha \text{ causes } \langle \gamma, \epsilon \rangle,$$

$$\alpha \text{ causes } \langle \beta, \epsilon', \gamma \rangle \vee \alpha \text{ causes } \langle \beta, \epsilon', \epsilon \rangle \vee \alpha \text{ causes } \langle \beta, \epsilon', \gamma, \epsilon \rangle,$$

$$\text{and the three implications } \alpha \rightarrow \beta, \beta \rightarrow (\gamma \vee \epsilon) \text{ and } \alpha \rightarrow \epsilon'.$$

We get the expected result $\alpha \rightarrow \gamma$ (from the four implications). However the causal formula $\varphi = \alpha \text{ causes } \langle \gamma \rangle$ is not a consequence of CT_1 .

There are two reasons for this: The partial disconnection between causal formulas and classical formulas on one hand. The restrictions put on causal formulas which prevent negations in causal atoms on the other hand.

Here is a simple modification which will give the formula φ . Let us separate the W part of any causal theory CT into a “*definitional part*” WD and a remaining part. The semantics is modified by replacing the inference \models in (6) of Definition 1 by \models_{WD} (inference in WD), $X \models_{WD} Y$ meaning $WD \cup \{X\} \models Y$. In CT_1 here, $\epsilon' \rightarrow \neg\epsilon$ would be put in WD , and with this modified causal inference \Vdash_D we get $CT_1 \Vdash_D \alpha \text{ causes } \langle \gamma \rangle$, in accordance with a natural expectation. The proof system in § 5 should then been modified by adding a rule generalizing to the case of not single effects the new behavior of this example, namely $CT_1 \rightarrow \alpha \text{ causes } \langle \gamma \rangle$.

Notice that this new semantics would not modify the behavior in case (7) of Remark 3 in § 4.7. One reason is that this would violate the property given in §4.8. Another way to see this is that putting $\delta \rightarrow \alpha \vee \beta$ in WD would not modify S_δ . If we wanted to get the conclusion $\delta \text{ causes } \langle \gamma \rangle$ in case (7) (which is not a desirable feature in our opinion), a more serious modification of the semantics, in the lines of condition (EE) in Remark 4 in §4.7, would be necessary.

7 Inferring Causal Explanations

There are two ways to reason from causes and effects. One is just deduction: From what is known to be true and what is known to cause what, infer what else is true. The other is abduction: From what is observed and what is known to cause what, infer what could explain the current facts. Notice that neither task is about discovering causal relationships: these are supposed to be already available and are simply used to infer deductive/abductive conclusions.

The above account of causes being rather strict, it can serve as a basis for both tasks.

As for deduction, it corresponds directly to causal inference \Vdash : the premises consist of facts and causal statements and the conclusions are statements expected to be true by virtue of the premises. typically, $(\alpha \vee \beta) \wedge \delta$ together with $\delta \text{ causes } \langle \gamma \rangle$ make $(\alpha \vee \beta) \wedge \gamma \wedge \delta$ to be deduced.

As for abduction, it works typically as follows: Consider the information stating that δ causes either α or β or γ . Consider further that *observation* β happens to be the case (it is a fact). Then, δ might explain that β has taken place. Hence the next definition:

Given a causal theory CT , δ is a causal explanation for β if
 $CT \Vdash \delta \text{ causes } \langle \gamma_1, \dots, \gamma_{i-1}, \beta, \gamma_{i+1}, \dots, \gamma_n \rangle$.

The explanation relationship does not propagate through equivalence (Remark 5).

It must be reminded that since *possible explanations* are inferred, there is no guarantee of inferring only the “right” explanation (if any). Most of the time, the *available* causal information is anyway not enough to determine what the right explanation is and a logic is not meant to go beyond the premises it is applied to.

The relations as defined here can be considered as too strict, and in practice they should be augmented by considering some “definition formulas” WD as explained in § 6.2, which would extend the range of application of the formalism.

8 Causal Relation versus Predictive Inference

Some logics for causal reasoning (e.g., [4, 7]) satisfy apparently much more properties than the formalism presented here. However, any comparison should keep in mind that here a new specific kind of “causal formulas” has been introduced by the way of the “causal atoms”. These causal atoms are not real atoms, as already remarked, since (1) they are physically made of smaller atoms, and (2) new “causal atoms” can be inferred from sets of causal atoms, as shown in § 5, and even “greater” causal atoms than those already present can be inferred, as shown in points (1) and (5) of the proof system (§ 5). However, they are “atoms” in a weak sense, which explains the relatively small number of properties allowing to derive new causal atoms.

This explains why, when making a comparison with most of the literature on the subject, the *predictive inference*, namely \Vdash , must be taken into account rather than the *causes* whose (presently known) properties are listed in §5.

Here are a few properties satisfied by the predictive inference \Vdash :

Material Implication $\{\alpha \text{ causes } \langle \gamma_1, \dots, \gamma_i \rangle\} \Vdash \alpha \rightarrow \gamma_1 \vee \dots \vee \gamma_i$

Strengthening $\{\alpha \text{ causes } \langle \gamma \rangle, \delta \rightarrow \alpha, \delta\} \Vdash \gamma$.

Right Weakening $\{\alpha \text{ causes } \langle \delta \rangle, \delta \rightarrow \gamma, \alpha\} \Vdash \gamma$.

Or (in antecedent) $\{\alpha \text{ causes } \langle \gamma \rangle, \beta \text{ causes } \langle \gamma \rangle, \delta \leftrightarrow \alpha \vee \beta, \delta\} \Vdash \gamma$.

Cut $\{\alpha \text{ causes } \langle \gamma \rangle, \beta \text{ causes } \langle \delta \rangle, \alpha \wedge \gamma \leftrightarrow \beta, \alpha\} \Vdash \delta$.

Since $\{\} \Vdash \top$ and $\{\alpha, \neg\alpha\} \Vdash \perp$ obviously hold, it looks like the predictive inference relation satisfies all the postulates of a *causal production relation* as defined in [4] and all the properties of a *disjunctive causal relation* [3]. It must be pointed out that our hypotheses are expressed here in terms of causal formulas while our conclusions pertain to classical logic.

Since the “disjunctive case” as considered e.g. in [3] does not really go “inside the disjunctive effects”, formalisms such as those from [3, 7] will not have the same behavior as ours when it comes to abduction. The reason is that inference in our logic strictly conforms with the causal chains which can effectively be obtained. If γ_1 happens to be inferred by means of the causal formula $\delta \text{ causes } \langle \gamma_1, \gamma_2 \rangle$ under the assumption δ (in symbols, $CT \Vdash \delta \text{ causes } \langle \gamma_1, \gamma_2 \rangle$ and $CT \cup \{\delta\} \Vdash \gamma_1$) then δ becomes an abductive conclusion but it would not be so if γ_1 were to be true for another reason (a purely deductive one). This feature

seems important for a correct treatment of disjunctive effects when dealing with abduction.

9 Perspectives and Conclusion

We have provided a logical framework intended to formalize causal relations, allowing predictive and abductive reasoning. Classical propositional logic has been extended by new causal formulas describing causal relations as they are known by the user. These causal formulas follow a semantics which has been tailored in order to get the expected conclusions, and no more. Also, these formulas admit only propositional atoms as premises and only set of such atoms (intending to model disjunctive effects) as conclusions. This is to keep the definitions simple enough.

Restricting the arguments of the causal operators to propositional atomic formulas is unsatisfactory. We have evoked two ways in order to overcome this problem: (1) Considering “definitional formulas”, which take a key role in the definition of the semantics, as explained in § 6.2. (2) Adopt a condition (such as (EE) in Remark 4 in §4.7), linking the causal configurations of formulas which are equivalent in W , which would extend even more the range of the predictive inference.

Yet, much remains to be done to extend the above logic to enjoy arbitrary formulas as arguments of the causal operators. Perhaps the main difficulty lies in the following incompatibility. Presumably, δ being a cause for α should not lead to the conclusion that δ is a cause for $\alpha \vee \beta$. However, δ being a cause for $\alpha \wedge \beta$ should entail that δ is a cause for α . Thus, δ being a cause for $(\alpha \vee \beta) \wedge \alpha$ should entail that δ is a cause for $\alpha \vee \beta$. As $(\alpha \vee \beta) \wedge \alpha$ is logically equivalent with α , it follows that δ being a cause for α would entail that δ is a cause for $\alpha \vee \beta$ – unless logically equivalent effects are not taken to share the same causes. Such a requirement is technically possible (as in the above logic) but is more problematic when arbitrary formulas occur as arguments of the causal operators: The statement that δ is a cause for $\alpha \leftrightarrow \neg\beta$ would fail to be equivalent to the statement that δ is a cause for $\beta \leftrightarrow \neg\alpha$ (similarly, δ causing $(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$ would not be equivalent with δ causing $\alpha \wedge (\beta \vee \gamma)$, and so on).

Another direction for generalization is to alleviate the constraint that a cause always brings about the effect. E.g., taking “Too much light causes blindness” to be true even though there would be some possibility that certain circumstances may tolerate too much light not to lead to blindness. A technical solution would be to introduce a constant \star in the language to stand for a “ghost” effect. In order to encode the example just given about light and blindness, l causes $\langle b, \star \rangle$ would do whereas l causes $\langle b \rangle$ would not hold. Special rules should obviously govern \star , so that δ causes $\langle \gamma_1, \dots, \gamma_n \rangle$ be consistent with $\neg(\delta$ causes $\langle \gamma_1, \dots, \gamma_n, \star \rangle)$ and δ causes $\langle \gamma_1, \dots, \gamma_n, \star \rangle$ be consistent with $\neg(\delta$ causes $\langle \gamma_1, \dots, \gamma_n \rangle)$.

We have presented the basis for developing a logic taking causal relations into consideration. Once the pertinent causal relations are known, together with

some background knowledge, the aim is to deduce some conclusions, or to abduce some hypotheses. Here, our goal was to define a set of basic incontestable rules. Then, real systems should be built upon this kernel, by adding some “ornament”. This is the place where notions such as *strong* or *definitional knowledge* (“is-a rules” in particular), and some *weaker knowledge* should be introduced. Also, in order to facilitate tasks such as diagnosis, notions of *observable formulas* and *abducible formulas* should be considered. We consider that the basic rules cannot be extended significantly without losing their adequacy with the nature of causation rules, and that real problems can be solved by taking care of the variety of the kinds of informations at hand. Also, some non-monotonic methods could be provided, e. g. by using the “ghost effect” evoked above.

Acknowledgments

It is a pleasure for the authors to thank the referees for their helpful and constructive comments.

References

1. Besnard Ph. and Cordier M.-O. Inferring Causal Explanations. In A. Hunter and S. Parsons (editors), *ECSQARU-99*, pp. 55–67, LNAI Vol. 1638, Springer, 1999.
2. Bell J. Causation and Causal Conditionals. In D. Dubois, C. Welty and M.-A. Williams (editors), *KR-04*, pp. 2–11, AAAI Press, 2004.
3. Bochman A. On Disjunctive Causal Inference and Indeterminism. In G. Brewka and P. Peppas (editors), *IJCAI Workshop: NRAC-03*, pp. 45–50, 2003.
4. Bochman A., A Causal Theory of Abduction. *Common Sense* 2005 pp. 33–38, 2005
5. Collins J., Hall N. and Paul L. A. (editors), *Causation and Counterfactuals*, MIT Press, 2004.
6. Chellas B. F. *Modal Logic*. Cambridge University Press, 1980.
7. Giunchiglia E., Lee J., Lifschitz V., McCain N., Turner H. Nonmonotonic Causal Theories. *Artificial Intelligence* 153(1–2):49–104, 2004.
8. Hall N. Causation and the Price of Transitivity. *Journal of Philosophy* 97:198–222, 2000.
9. Halpern J. and Pearl J. Causes and Explanations: A structural-model approach. Part I: Causes. In J. S. Breese and D. Koller (eds), *UAI-01*, pp. 194–202, Morgan Kaufmann, 2001.
10. Halpern J. and Pearl J. Causes and Explanations: A Structural-Model Approach - Part II: Explanations. In B. Nebel (ed), *IJCAI-01*, pp. 27–34, Morgan Kaufmann, 2001.
11. Lewis D. Causation. *Journal of Philosophy* 70:556–567, 1973.
12. Lewis D. *Counterfactuals*. Blackwell, 1973.
13. Lewis D. Causation as Influence. *Journal of Philosophy* 97:182–197, 2000.
14. Mellor D. H. *The Facts of Causation*. Routledge, 1995.
15. Nute D. *Topics in Conditional Logic*. Reidel, 1980.
16. Pearl J. *Causality*. Cambridge University Press, 2000.
17. Shafer G. *The Art of Causal Conjecture*. MIT Press, 1996.
18. Shafer G. Causal Logic. In H. Prade (ed), *ECAI-98*, pp. 711–720, Wiley, 1998.
19. Yablo S. Advertisement for a Sketch of an Outline of a Proto-Theory of Causation. In [5].