Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
48 (2008) 752–765
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

# Predicting causality ascriptions from background knowledge: model and experimental validation ☆

Jean-François Bonnefon [a,*], Rui Da Silva Neves [a], Didier Dubois [b], Henri Prade [b]

[a] *Université de Toulouse, CLLE, Maison de la Recherche, 5, allées A. Machado, 31058 Toulouse Cedex 9, France*
[b] *IRIT-CNRS, 118 route de Narbonne, 31062 Toulouse Cedex, France*

Available online 27 July 2007

**Abstract**

A model is defined that predicts an agent's ascriptions of causality (and related notions of facilitation and justification) between two events in a chain, based on background knowledge about the normal course of the world. Background knowledge is represented by non-monotonic consequence relations. This enables the model to handle situations of poor information, where background knowledge is not accurate enough to be represented in, e.g., structural equations. Tentative properties of causality ascriptions are discussed, and the conditions under which they hold are identified (preference for abnormal factors, transitivity, coherence with logical entailment, and stability with respect to disjunction and conjunction). Empirical data are reported to support the psychological plausibility of our basic definitions.
© 2007 Elsevier Inc. All rights reserved.

## 1. Introduction

The problem of causal ascription that we will consider in this article needs to be carefully distinguished from two other causality problems more commonly studied in Artificial Intelligence (AI), namely, diagnosis and the simulation of dynamical systems. Note, however, that making a distinction between different problems where causality is involved does not presuppose any opinion about whether causality is a unique notion – it only suggests that different problems can emphasize different aspects of a possibly unique notion .

Diagnosis problems are basically a matter of abduction: One takes advantage of the knowledge of some causal links to infer the most plausible causes of an observed event [1].[1] In this setting, causality relations are often modelled by conditional probabilities Pr(effect|cause). Nevertheless, Bayesian networks [3] that represent a joint probability distribution by means of a directed graph do not necessarily reflect causal links between their nodes, because different graphical representations can be obtained depending on the ordering

☆ Supported by a grant from the Agence Nationale pour la Recherche, project number NT05-3-44479.
* Corresponding author. Tel.: +33 561503546; fax: +33 561503589.
*E-mail address:* bonnefon@univ-tlse2.fr (J.-F. Bonnefon).
[1] However, model-based diagnosis is rather a matter of inconsistency checking, finding contradictions between good behavior assumptions and current observations of a system [2], and does not refer to ideas of causality.

0888-613X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2007.07.003

Author's personal copy



in which variables are considered [4] – a problem that has been tackled by the notion of *intervention* [5], to which we will come back at the end of this article.

Dynamical systems are modelled in AI in terms of qualitative physics [6], or by means of logics of action [7–9]. Material implication being inappropriate to represent a causal link, the latter approaches define a 'causal rule' as 'if action $A$ has been executed then there exists a cause for $B$,' where 'there exists a cause for $B$' is expressed by means of a modal operator. Such causal theories are generally non-monotonic, leaving room for abnormality. Indeed, the behavior of these causal theories tends to minimize unexplained (uncaused) abnormality [10]. Our approach will give a central role to the notion of abnormality in the detection of causal relations, in contrast with the aforementioned causal theories [7–10], in which abnormality is merely minimized, and only denotes the lack of an explanation. Furthermore, these approaches only establish the *existence* of causes, but do not identify causes as such.

The problem of causal ascription discussed in this paper is not one of abductive diagnosis (neither does it deal with the qualitative simulation of dynamical systems, nor with the problem of describing changes caused by the execution of actions, nor with what does not change when actions are performed). Our problem is to infer an unknown causal relation from two known events and some background non-causal knowledge.[2] We are concerned here with the ascription of causal relations within a sequence of reported events, that is, the detection of pairs of events that can be considered as related by a causality relation, under the normal course of things. In some sense, our work is reminiscent of the 'causal logic' of Shafer [11], which provides a logical setting that aims at describing the possible relations of concomitance between events when an action takes place. However, Shafer's logic does not leave room for abnormality. This notion is central in our approach, as our view of normal causality directly relates to relations of qualitative independence explored in possibility theory [12] – causality and independence being somewhat antagonistic notions.

Models of causal ascriptions presuppose a representation of the underlying causality-ascribing agent's knowledge. Unlike standard diagnosis problems, causality ascription is a problem of describing as 'causal' the link between two observed events in a sequence. The first step in modelling causal ascription is to define causality in the language chosen for the underlying representation of knowledge. In this article, we define and discuss a model of causal ascription that represents knowledge by means of non-monotonic consequence relations obeying the rules of System P.[3] Indeed, agents often possess poor knowledge about the world, under the form of default rules. Clearly, this type of background knowledge is less accurate than, e.g., structural equations. It is nevertheless appropriate to predict causal ascriptions in situations of restricted knowledge. We first present the logical language we will use to represent background knowledge. We then define our main notion of causality and establish some formal properties of the model. Next, we introduce a new notion (facilitation), which is less committing than causality, in terms of the beliefs required from the agent. Empirical data on facilitation vs. causality ascriptions are reported to support the distinction between these two notions. Finally, we relate our model to other works on causality in AI, and distinguish the notion of epistemic justification from that of causality.

## 2. Ascribing causality

An agent capable of acknowledging a causality link between two reported events must possess some background knowledge allowing the recognition of normal patterns of occurrence in a set of reported facts. Indeed, some recent work in AI has emphasized the role of norms (in both the deontic and non-deontic meanings of the term) for ascribing causal links in car accident reports [17,18]. Hence, a prerequisite for a proper definition of causality ascription is a language for describing the agent's generic knowledge. This knowledge is generally qualitative in nature, but should tolerate exceptional situations, since an agent is often capable of distinguishing normal courses of things from abnormal ones.

In the following definitions, $A$, $B$, $C$, and $F$ are events modelling either actions or descriptions of states of affairs. Notations do not discriminate between actions and descriptions, since this distinction does not yet play a role in the model. Events are time-stamped (e.g., $A_t$) when they are reported facts.

[2] In contrast, abductive diagnosis amounts to inferring an unknown event (the cause) from a known event (the effect) and a known causal relation.

[3] This model was first advocated in two workshop papers [13,14]. The present paper is an expanded version of [15,16].

Author's personal copy



### 2.1. Modeling background knowledge

An agent is supposed to have observed or learned of a sequence of events, e.g. $\neg B_t$, $A_t$, $B_{t+k}$. This expresses that $B$ was false at time $t$, when $A$ took place, and that $B$ became true afterwards ($t+k$ denotes a time point after $t$). There is no uncertainty about these events.

Moreover, the agent maintains a knowledge base made of conditional statements of the form 'in context $C$, if $A$ takes place then $B$ is generally true afterwards', or 'in context $C$, $B$ is generally true'. These will be denoted by $A_t \wedge C_t \mid\!\sim B_{t+k}$, and by $C_t \mid\!\sim B_t$, respectively. (Time indices will be omitted when there is no risk of confusion.) The conditional beliefs of an agent with respect to $B$ when $A$ takes place or not in context $C$ can take three forms: (i) If $A$ takes place $B$ is generally true afterwards: $A_t \wedge C_t \mid\!\sim B_{t+k}$; (ii) If $A$ takes place $B$ is generally false afterwards: $A_t \wedge C_t \mid\!\sim \neg B_{t+k}$; (iii) If $A$ takes place, one cannot say whether $B$ is generally true or false afterwards. In this case, neither $A_t \wedge C_t \mid\!\sim B_{t+k}$ nor $A_t \wedge C_t \mid\!\sim \neg B_{t+k}$ holds. The fact that an agent cannot assert $A \mid\!\sim B$ is denoted by $A \not\mid\!\sim B$.

We assume that the non-monotonic consequence relation $\mid\!\sim$ satisfies the requirements of 'System P' [19]; namely, $\mid\!\sim$ is reflexive and the following postulates and characteristic properties hold ($\models$ denotes classical logical entailment):

- *Right Weakening*: $E \mid\!\sim F$ and $F \models G$ imply $E \mid\!\sim G$
- *Left AND*: $E \mid\!\sim F$ and $E \mid\!\sim G$ imply $E \mid\!\sim F \wedge G$
- *Right OR*: $E \mid\!\sim G$ and $F \mid\!\sim G$ imply $E \vee F \mid\!\sim G$
- *Cautious monotony*: $E \mid\!\sim F$ and $E \mid\!\sim G$ imply $E \wedge F \mid\!\sim G$
- *Cut*: $E \mid\!\sim F$ and $E \wedge F \mid\!\sim G$ imply $E \mid\!\sim G$

As we are describing propositions as events (or subsets of possible worlds), and not as well-formed formulas of propositional logic, we do not need any syntax independence axiom. Right Weakening and Left AND together imply that the set of beliefs of the agent is deductively closed. Right OR avoids the need for reasoning by cases. Cut follows from the other axioms. Cautious Monotony and Cut replace Monotony and Transitivity properties of classical inference so as to lay bare the possibility of exceptional situations, and make this setting maximally cautious.

System P enjoys only one half of the deduction theorem:

- *HD*: $E \wedge F \mid\!\sim G$ implies $E \mid\!\sim \neg F \vee G$.

In addition, we shall sometimes assume the property of Rational Monotony [20], a strong version of Cautious Monotony involving an explicit handling of operator $\not\mid\!\sim$ as a means of reasoning about ignorance:

- *Rational Monotony*: $E \not\mid\!\sim \neg F$ and $E \mid\!\sim G$ imply $E \wedge F \mid\!\sim G$

Empirical studies repeatedly demonstrated [21–25] that System P and Rational Monotony provide a psychologically plausible representation of background knowledge and default inference. Arguments for using non-monotonic logics in modeling causal reasoning were also discussed in the cognitive science literature [26]. Finally, system P is known to be a qualitative variant of probabilistic reasoning, since the axioms of system P hold for infinitesimal conditional probabilities [27,3]. These axioms hold as well for a special kind of standard probabilities, namely, big-stepped probabilities [28,29]. There also exists a possibilistic semantic for System P, which holds for any kind of possibility measure [30]. Moreover, adding Rational Monotony comes down to reasoning with a single possibility distribution in possibilistic logic [30].

### 2.2. A definition of causality ascription

Assume that in a given context, described by the situation where the agent knows that $C$ holds, the occurrence of event $B$ is considered to be exceptional (i.e., $C \mid\!\sim \neg B$). Assume now that for some event $A$, it is part of the agent's background knowledge that $A \wedge C \mid\!\sim B$. If both conditions are satisfied, we will say that in context $C$, $A$ is perceived to be the *cause* of $B$ when an agent learns that $B$ was false when $A$ was reported, then $B$ became true.

Author's personal copy



**Definition 1** (*Causality ascription*). Let us assume that an agent learns of the sequence $\neg B_t$, $A_t$, $B_{t+k}$. Let us call $C_t$ (the *context*) the conjunction of all other facts known by, or reported to, the agent at time $t$. If the agent possesses the pieces of default knowledge that $C \mid\!\sim \neg B$ and $A \wedge C \mid\!\sim B$, the agent will perceive $A_t$ to be the cause of $B_{t+k}$ in context $C_t$, denoted $C_t : A_t \Rightarrow_{\mathrm{ca}} B_{t+k}$.

In the above definition $A$ can stand for any compound reported fact such as $A' \wedge A''$. Note that our formal framework for causality ascription is based on the use of system P: that is, it only uses pieces of background knowledge featuring $\mid\!\sim$. However, let us stress that the causality relation $C : A \Rightarrow_{\mathrm{ca}} B$ is distinct from the non-monotonic consequence relation $\mid\!\sim$. This feature departs from works viewing causal relations as particular non-monotonic consequence relations (e.g., [31]).

**Example 2** (*Driving while intoxicated*). When driving, one has generally no accident, $Drive \mid\!\sim \neg Accident$. This is no longer true when driving fast while drunk, which normally leads to an accident, $Drive \wedge Fast \wedge Drunk \mid\!\sim Accident$. Suppose now that an accident took place after the driver drove fast while being drunk. $Fast \wedge Drunk$ will be perceived as the cause of the accident.

In our model, $C \mid\!\sim \neg B$ and $A \wedge C \mid\!\sim B$ must be understood as pieces of default knowledge used by the agent to interpret the chain of reported facts $\neg B_t$ (in context $C$), $A_t$, $B_{t+k}$. An interesting situation arises when, in context $C$, an agent learns of the sequence $\neg B_t$, $A_t$, and $B_{t+k}$, while it believes that $\neg B_t \wedge C \mid\!\sim \neg B_{t+k}$, and that $A_t \wedge C \mid\!\sim \neg B_{t+k}$. Then the agent cannot consider that $C : A_t \Rightarrow_{\mathrm{ca}} B_{t+k}$, and it may suspect some fact went unreported: finding about it would amount to a diagnosis problem. In contrast, when an agent believes that $C \mid\!\sim \neg B$ and $A \wedge C \mid\!\sim B$, and learns of the sequence of events $\neg B_t$, $A_t$, and $\neg B_{t+k}$, the agent would conclude that $A_t$ failed to produce its normal consequence, for unknown reasons.

The introduction of the parameter $k$ in Definition 1 implicitly refers to the delay usually required for $A$ to produce its effect, (namely, to make $B$ happen). This would suggest a further condition for the ascription of causality, namely, that the value of $k$ be consistent with the agent's beliefs about such a delay when $A$ takes place. In the rest of this article, we will assume this condition to be satisfied.

## 3. Properties of causal ascriptions

In this section, properties of our definition of causality are reviewed. The validity of the results below presupposes the setting of system P only.

### 3.1. Impossibility of mutual causality

**Proposition 3.** *From the minimal sequence of events required for $C : A \Rightarrow_{ca} B$, it is impossible to believe $C : B \Rightarrow_{ca} A$.*

**Proof.** The ascription by an agent that $C : A \Rightarrow_{\mathrm{ca}} B$ requires (in addition to particular beliefs of the agent) that the minimal sequence $\{\neg B_t, A_t, B_{t+k}\}$ has taken place. This is clearly incompatible with the simultaneous ascription that $C : B \Rightarrow_{\mathrm{ca}} A$, since there is obviously no $t'$, $k'$ such that $\{\neg B_t, A_t, B_{t+k}\}$ can be reconciliated with $\{\neg A_{t'}, B_{t'}, A_{t'+k'}\}$. $\square$

Note that Proposition 3 does not exclude that, in turn, $A$ is perceived to cause $B$, then, later, $B$ is perceived to cause $A$, in an oscillatory-like fashion.

### 3.2. Preference for abnormal causes

Psychologists show that abnormal conditions are more likely to be selected by human agents as the cause of an event [32] and all the more so if this event is itself abnormal [33] (see also [34] in the area of legal philosophy). Our model reflects this preference: Only what is abnormal in a given context can be perceived as causing a change in the normal course of things in this context.

Author's personal copy



**Proposition 4.** *If* $C : A \Rightarrow_{ca} B$ *then* $C \mid\!\sim \neg A$.

**Proof.** If $C : A \Rightarrow_{\mathrm{ca}} B$, it holds that $C \mid\!\sim \neg B$, $C \wedge A \mid\!\sim B$. Using HD, $C \wedge A \mid\!\sim B$ implies $C \mid\!\sim \neg A \vee B$. Using AND, we get $C \mid\!\sim (\neg A \vee B) \wedge \neg B$, that is $C \mid\!\sim \neg A \wedge \neg B$, and by RW, $C \mid\!\sim \neg A$. □

**Example 5** (*The unreasonable driver*). Let us imagine an agent who believes it is normal to be drunk in the context of driving (*Drive* $\mid\!\sim$ *Drunk*). This agent may think that it is exceptional to have an accident when driving (*Drive* $\mid\!\sim \neg$*Accident*). In that case, the agent cannot believe that accidents are exceptional as well when driving while drunk: *Drive* $\wedge$ *Drunk* $\mid\!\sim \neg$*Accident*. As a consequence, when learning that someone got drunk, drove his car, and had an accident, this agent will not consider that *Drive* : *Drunk* $\Rightarrow_{\mathrm{ca}}$*Accident*.

A notable and straightforward consequence of Proposition 4 is that an agent who perceives $A$ to be the cause of $B$ in context $C$ will also assume, if not told otherwise, that $A$ was false before it became true at time $t$ – provided that context $C$ is assumed to have been stable for some time before $t$.

### 3.3. Transitivity

Although many models of causation consider that causality should be transitive (e.g., [35]), Definition 1 does not grant general transitivity to the causal ascription $\Rightarrow_{\mathrm{ca}}$ – only a restricted form of transitivity holds. Note that it is an open question whether transitivity is desirable in the specific case of causality ascription.

If $C : A \Rightarrow_{\mathrm{ca}} B$ and $C : B \Rightarrow_{\mathrm{ca}} D$, it does not always follow that $C : A \Rightarrow_{\mathrm{ca}} D$. Formally: $C \mid\!\sim \neg B$ and $A \wedge C \mid\!\sim B$ and $C \mid\!\sim \neg D$ and $B \wedge C \mid\!\sim D$ do not entail $C \mid\!\sim \neg D$ and $A \wedge C \mid\!\sim D$, because $\mid\!\sim$ itself is not transitive. Although $\Rightarrow_{\mathrm{ca}}$ is not generally transitive, it becomes so in one particular case.

**Proposition 6.** *If* $C : A \Rightarrow_{\mathrm{ca}} B$, $C : B \Rightarrow_{\mathrm{ca}} D$, $B \wedge C \mid\!\sim A$, *and* $D$ *reportedly took place after* $A$, *then* $C : A \Rightarrow_{ca} D$.

**Proof.** From the definition of $C : B \Rightarrow_{\mathrm{ca}} D$, it holds that $B \wedge C \mid\!\sim D$. From $B \wedge C \mid\!\sim A$ and $B \wedge C \mid\!\sim D$, applying Cautious Monotony yields $A \wedge B \wedge C \mid\!\sim D$, which together with $A \wedge C \mid\!\sim B$ (from the definition of $C : A \Rightarrow_{\mathrm{ca}} B$) yields by Cut $A \wedge C \mid\!\sim D$; since it holds from the definition of $C : B \Rightarrow_{\mathrm{ca}} D$ that $C \mid\!\sim \neg D$, the two parts of the definition of $C : A \Rightarrow_{\mathrm{ca}} D$ that involve background knowledge are satisfied. Furthermore, $C : A \Rightarrow_{\mathrm{ca}} B$ requires a sequence $\{A_t, \neg B_t, B_{t+k}\}$ and $C : B \Rightarrow_{\mathrm{ca}} D$ requires a sequence $\{B_{t'}, \neg D_{t'}, D_{t'+k'}\}$. From $C \mid\!\sim \neg D$, it holds that $\neg D_t$ (while it holds that $D_{t'+k'}$, it cannot be the case that $t = t' + k'$ from the condition that $D$ reportedly took place after $A$, that is, $t < t' + k'$). Hence the sequence $\{A_t, \neg D_t, D_{t'+k'}\}$ is valid, as required by $C : A \Rightarrow_{\mathrm{ca}} D$. □

**Example 7** (*Mud on the plates*). Driving back from the countryside, you get a fine because your plates are muddy, *Drive* : *Mud* $\Rightarrow_{\mathrm{ca}}$*Fine*. Let us assume that you perceive your driving to the countryside as the cause for the plates to be muddy, *Drive* : *Countryside* $\Rightarrow_{\mathrm{ca}}$*Mud*. Transitivity will apply (and yield *Drive*:*Countryside* $\Rightarrow_{\mathrm{ca}}$*Fine*) as soon as it holds hold that *Mud* $\wedge$ *Drive* $\mid\!\sim$ *Countryside*: If mud on your plates usually means that you went to the countryside, then the trip can be considered the cause of the fine. If the presence of mud on your plates does not allow to infer that you went to the countryside (perhaps you also regularly drive through muddy streets where you live), then transitivity is no longer warranted; you will only consider that the mud caused the fine, not that the trip did.

Note also that the restricted transitivity property expressed in Proposition 6 agrees well with the fact that reports often identify actions with their consequences, when the latter are prototypically diagnostic of the former. For example, one may either say that a fast driver had an accident because he had been drinking (action) or because he was inebriated (consequence). Indeed, in this situation, one can assume that *Inebriated* $\mid\!\sim$ *Drinking*.

### 3.4. Entailment and causality ascriptions

Classical entailment $\models$ does not preserve $\Rightarrow_{\mathrm{ca}}$. If $C : A \Rightarrow_{\mathrm{ca}} B$ and $B \models B'$, one cannot say that $C : A \Rightarrow_{\mathrm{ca}} B'$. Indeed, while $A \wedge C \mid\!\sim B'$ follows by right weakening [19] from $A \wedge C \mid\!\sim B$, it is not generally true that $C \mid\!\sim \neg B'$, given that $C \mid\!\sim \neg B$.

Author's personal copy



Besides, according to Definition 1, if $A' \models A$, the fact that $C : A \Rightarrow_{\mathrm{ca}} B$ does not entail that $C : A' \Rightarrow_{\mathrm{ca}} B$, since $C \mid\!\sim \neg B$ and $A \wedge C \mid\!\sim B$ do not entail $A' \wedge C \mid\!\sim B$ when $A' \models A$. This fact is due to the extreme cautiousness of System P. In contrast, assuming Rational Monotony would enable the latter inference insofar as $A \wedge C \mid\!\sim \neg A'$ does not hold, as shown in the following example.

**Example 8** (*Stone throwing*). An agent believes that a window shattered because a stone was thrown at it ($Window : Stone \Rightarrow_{\mathrm{ca}} Shatter$), based on its beliefs that $Window \mid\!\sim \neg Shatter$ and $Stone \wedge Window \mid\!\sim Shatter$. Using the Cautious Monotony of System P, it is not possible to predict that the agent would make a similar ascription if a *small* stone had been thrown (*SmallStone*), or if a *white* stone had been thrown (*WhiteStone*), or even if a *big* stone had been thrown (*BigStone*), although it holds that $SmallStone \models Stone$, $WhiteStone \models Stone$, and $BigStone \models Stone$. Adding Rational Monotony [20] to System P allows the ascriptions $Window{:}BigStone \Rightarrow_{\mathrm{ca}} Shatter$ and $Window : WhiteStone \Rightarrow_{\mathrm{ca}} Shatter$, but also $Window : SmallStone \Rightarrow_{\mathrm{ca}} Shatter$. To block this last ascription, it would be necessary that the agent has specific knowledge about the harmlessness of small stones, such as $Window \wedge Smallstone \mid\!\sim Shatter$ or even $Window \wedge Smallstone \mid\!\sim \neg Shatter$ (or if it were known that stones thrown at windows are generally not small).

### 3.5. Stability with respect to disjunction and conjunction

$\Rightarrow_{\mathrm{ca}}$ is stable with respect to disjunction, both on the right and on the left, and stable with respect to conjunction on the right. Such properties were laid bare in [12] in the setting of qualitative possibility theory. The following proposition shows their validity in system P.

**Proposition 9.** *The following properties hold*:

(1) *If* $C : A \Rightarrow_{ca} B$ *and* $C : A \Rightarrow_{ca} B'$, *then* $C : A \Rightarrow_{ca} B \wedge B'$.
(2) *If* $C : A \Rightarrow_{ca} B$ *and* $C : A \Rightarrow_{ca} B'$, *then* $C : A \Rightarrow_{ca} B \vee B'$.
(3) *If* $C : A \Rightarrow_{ca} B$ *and* $C : A' \Rightarrow_{ca} B$, *then* $C : A \vee A' \Rightarrow_{ca} B$.

**Proof.** Applying AND to the first part of the definitions of $C : A \Rightarrow_{\mathrm{ca}} B$ and $C : A \Rightarrow_{\mathrm{ca}} B'$, i.e., $C \mid\!\sim \neg B$ and $C \mid\!\sim \neg B'$, yields $C \mid\!\sim \neg B \wedge \neg B'$, which together with Right Weakening, yields $C \mid\!\sim \neg B \vee \neg B'$, and thus $C \mid\!\sim \neg(B \wedge B')$. Now, applying AND to the second part of the definitions of $C : A \Rightarrow_{\mathrm{ca}} B$ and $C : A \Rightarrow_{\mathrm{ca}} B'$, i.e., $A \wedge C \mid\!\sim B$ and $A \wedge C \mid\!\sim B'$, yields $A \wedge C \mid\!\sim B \wedge B'$. The definition of $C : A \Rightarrow_{\mathrm{ca}} B \wedge B'$ is thus satisfied, and Fact 1 is proved.

The second fact is proved similarly, just noticing that $C \mid\!\sim \neg B \wedge \neg B'$, and $A \wedge C \mid\!\sim B \vee B'$, obtained from $A \wedge C \mid\!\sim B \wedge B'$ by Right weakening, is exactly $C : A \Rightarrow_{\mathrm{ca}} B \vee B'$.

The proof of Fact 3 is obtained by separately applying OR to the first parts and the second parts of the definitions of $C : A \Rightarrow_{\mathrm{ca}} B$ and $C : A \Rightarrow_{\mathrm{ca}} B'$. □

$\Rightarrow_{\mathrm{ca}}$ is not stable with respect to conjunction on the left. If $C : A \Rightarrow_{\mathrm{ca}} B$ and $C : A' \Rightarrow_{\mathrm{ca}} B$, then it is not always the case that $C : A \wedge A' \Rightarrow_{\mathrm{ca}} B$.

Note that left conjunction of causal ascriptions would be pragmatically incongruous. If one believes that $C : A \Rightarrow_{\mathrm{ca}} B$ and that $C : A' \Rightarrow_{\mathrm{ca}} B$, it would be quite misleading to assert that $C : A \wedge A' \Rightarrow_{\mathrm{ca}} B$ from a conversational point of view, as it would suggest that *both* $A$ and $A'$ were needed to make $B$ happen. If indeed either one of $A$ and $A'$ are perceived as sufficient for having caused $B$, it is more cooperative to assert $C : A \vee A' \Rightarrow_{\mathrm{ca}} B$ than $C : A \wedge A' \Rightarrow_{\mathrm{ca}} B$.

**Example 10** (*Busy professors*). Suppose that professors in your department seldom show up early at the office ($Prof \mid\!\sim \neg Early$). However, they generally do so when they have many student papers to mark ($Prof \wedge Mark \mid\!\sim Early$), and also when they have a grant proposal to write ($Prof \wedge Grant \mid\!\sim Early$). Today, a colleague is showing up early, and you know she has many papers to mark and a grant proposal to write. You are ready to say that the papers caused her to come early ($Prof : Mark \Rightarrow_{\mathrm{ca}} Early$), and also to say that the grant proposal caused her to come early ($Prof : Grant \Rightarrow_{\mathrm{ca}} Early$). Would you find it appropriate to say that the

Author's personal copy



cause of her showing up early was the papers *and* the grant? Probably not, as it would give the misleading impression that one and only one of those would have been insufficient for her to come in early.

More formally, the failure of left conjunction for causal ascriptions is once again due to the cautiousness of System P; for $C : A \wedge A' \Rightarrow_{\mathrm{ca}} B$ to hold, it is necessary that $C \wedge A \mid\!\sim A'$ or, alternatively, that $C \wedge A' \mid\!\sim A$. Then Cautious Monotony will yield $A \wedge A' \wedge C \mid\!\sim B$. Rational Monotony can soften this constraint and make it enough that $C \wedge A \not\mid\!\sim \neg A'$ or $C \wedge A' \not\mid\!\sim \neg A$. Thus, left conjunction will fail when it is abnormal to observe $A$ and $A'$ at the same time, for example when $A \mid\!\sim \neg A'$. Note that it may also happen that $A \wedge A' \wedge C \mid\!\sim \neg B$ without creating any inconsistency.

**Example 11** (*Busy professors*, *continued*). Suppose that in your department, it is very uncommon to have many papers to mark and a grant proposal to write on the same day, say, $Grant \mid\!\sim \neg Mark$. This might make it even more unlikely that you will say your colleague came in early because she had many papers to write *and* a grant proposal to write. If $Grant \mid\!\sim \neg Mark$, it is impossible to feel sure that $Prof : Mark \wedge Grant \Rightarrow_{\mathrm{ca}} Early$. For example, it might be the case that faced with such an exceptional workload, professors usually prefer working at home all day rather than coming to the office and be distracted. If $Grant \mid\!\sim Mark$ does not hold, then it is also impossible to come to the conclusion that $Prof : Mark \wedge Grant \Rightarrow_{\mathrm{ca}} Early$. In that situation, when a professor does anyway show up early when she has papers to mark and a proposal to write, one might look for another explanation: maybe she has an important meeting early in the morning, that she could not reschedule even though she wanted to stay home and work?

## 4. Facilitation: definition and experimentation

Causality is quite a strong notion, and one might suspect that some notion not quite as strong may also make sense in some situations. For instance, suppose some fact is generally believed to hold in a certain context, and the agent does not believe in that fact any longer in some restricted context, without necessarily believing its contrary. In this case the agent will be more cautious in its causal interpretation of the sequence of events. In this section, we model a companion relation to causality, that we call facilitation. Modelling this relation requires to complement System P with Rational Monotony. Furthermore, the distinction between causality and the new notion of facilitation is sanctioned, as we will see, by the results of two experiments.

### 4.1. A variant of causality: facilitation

Assume that in a given context $C$, the occurrence of event $B$ is known to be exceptional (i.e., $C \mid\!\sim \neg B$), and that indeed $\neg B$ is observed. Assume now that, further on, a fact $F$ is reported along with $B$, which becomes true. If $F$ is such that the agent believes neither $F \wedge C \mid\!\sim \neg B$ nor $F \wedge C \mid\!\sim B$ (respectively denoted $F \wedge C \not\mid\!\sim \neg B$ and $F \wedge C \not\mid\!\sim B$), we will say that in context $C$, $F$ alone is perceived to have *facilitated* the occurrence of $B$ (denoted $C : F \Rightarrow_{\mathrm{fa}} B$), since in some sense the occurrence of $F$ makes the occurrence of $B$ unsurprising (but not expected) to the agent.

**Definition 12** (*Facilitation ascription*). Let us assume that an agent learns of the sequence $\neg B_t$, $F_t$, $B_{t+k}$. Let us call $C_t$ (the *context*) the conjunction of all other facts known by or reported to the agent at time $t$. If the agent possesses the piece of default knowledge that $C \mid\!\sim \neg B$, but it holds for the agent that $F \wedge C \not\mid\!\sim \neg B$ and $F \wedge C \not\mid\!\sim B$, the agent will perceive $F_t$ as having facilitated the occurrence of $B_{t+k}$ in context $C_t$, denoted $C_t : F_t \Rightarrow_{\mathrm{fa}} B_{t+k}$.

**Example 13** (*Driving while intoxicated*, *again*). When driving, one has generally no accident, $Drive \mid\!\sim \neg Accident$. This is no longer true when driving while drunk, which is not as safe ($Drive \wedge Drunk \mid\!\sim \neg Accident$), even though it does not systematically or almost systematically generate accidents ($Drive \wedge Drunk \not\mid\!\sim Accident$). Suppose now that an accident took place, the driver being drunk. *Drunk* will only be judged as having *facilitated* the accident. In order to make a causality ascription, the agent needs a stronger piece of evidence, for instance, that not only the driver was drunk but that he drove fast, as accidents are much more likely to occur in this case.

Author's personal copy



Here, $F \wedge C \not\mid\!\sim \neg B$ stands as the absence of the piece of default knowledge $F \wedge C \mid\!\sim \neg B$ in the generic knowledge base of the agent. In fact, $\mid\!\sim$ could be understood in two different ways: Either $F \wedge C \not\mid\!\sim \neg B$ just means that $F \wedge C \mid\!\sim \neg B$ is not deducible from the agent's knowledge base, or it could mean that the agent really knows it is impossible to assert $F \wedge C \mid\!\sim \neg B$. The latter interpretation is stronger, and needs a more expressive language than the one of system P. The first interpretation is the right one here, because Rational Monotony means that the agent reasons in a monotonic way unless something that blocks monotonicity is believed, and not that the agent reasons in a monotonic way when it is aware not to know a certain conditional. The modelling of facilitation thus requires a conditional knowledge-base closed under System P and Rational Monotony, for instance obeying Rational Closure, i.e. the setting of possibility theory [30].

Note that Definition 12 is less committing than saying that $F$ 'prevents' $\neg B$ from persisting: $\not\mid\!\sim$ does not allow the jump from 'not having $\neg B$' to '$B$'. In Definition 1, the fact that $B$ is exceptional in context $C$ precludes the possibility for $C$ to be perceived as the cause of $B$ – but not the possibility that $B \models C$, i.e., that $C$ is a necessary condition of $B$. Thus, context can be a necessary condition of $B$ without being perceived as its cause.

An interesting situation arises when an agent only knows that $C \mid\!\sim \neg B$ and $F \wedge C \not\mid\!\sim \neg B$, and learns of the sequence of events $\neg B_t$ (in context $C$), $F_t$, $B_{t+k}$. Although this situation should lead the agent to judge that $C : F_t \Rightarrow_{\mathrm{fa}} B_{t+k}$, it may be tempting to judge that $C{:}F_t \Rightarrow_{\mathrm{ca}} B_{t+k}$, as long as no other potential cause reportedly took place.

Moreover, Proposition 4 has a counterpart for facilitation:

**Proposition 14.** *If* $C : F \Rightarrow_{fa} B$, *then* $C \mid\!\sim \neg F$.

**Proof.** Assume $C \not\mid\!\sim \neg F$. The Rational Monotony of $\mid\!\sim$ enforces $C \wedge F \not\mid\!\sim \neg B$ from $C \mid\!\sim \neg F$ and $C \not\mid\!\sim \neg B$; thus, it cannot be the case that $C : F \Rightarrow_{\mathrm{fa}} B$ since the facilitation relation requires that $C \wedge F \mid\!\sim \neg B$. □

The abnormality of a fact $F$ perceived as facilitating another fact $B$ can thus be established under the assumption of Rational Monotony.

**Example 15** (*The unreasonable driver is back*). Let us imagine an agent who believes it is normal to be drunk in the context of driving ($Drive \mid\!\sim Drunk$). This agent may think that it is exceptional to have an accident when driving ($Drive \mid\!\sim \neg Accident$). In that case, the agent cannot but believe that accidents are exceptional as well when driving while drunk: $Drive \wedge Drunk \mid\!\sim \neg Accident$. As a consequence, when learning that someone got drunk, drove his car, and had an accident, this agent will neither consider that $Drive : Drunk \Rightarrow_{\mathrm{ca}} Accident$ nor that $Drive : Drunk \Rightarrow_{\mathrm{fa}} Accident$.

There is no previous empirical evidence supporting the distinction we introduce between ascriptions of cause and facilitation. To check whether this distinction has intuitive appeal to lay reasoners, we conducted two experiments in which we presented participants with different sequences of events. We assessed their relevant background knowledge, from which we predicted the relations of cause and facilitation they should ascribe between the events in the sequence. We then compared these predictions to their actual ascriptions.

### 4.2. Experiment 1

#### 4.2.1. Methods

Participants were 46 undergraduate students, untrained in formal logic or in philosophy. Participants read the stories of three characters, and answered six questions after reading each story. In this section, we give a detailed presentation of the first story and the six questions that followed, and summarize the rest of the material. The first story read as follows:

> Benoît, who had never felt especially tired, recently took nightshifts at work, with a new boss who turned out to be very stressful. One month later, Benoît constantly feels very tired.

The first three questions were meant to check out each participant's background knowledge regarding the relations between nightshifts and feeling constantly tired, working under a stressful boss and feeling constantly tired, and the relation between nightshifts under a stressful boss and feeling constantly tired. For example, the first question read:

Author's personal copy



> What do you think is the most common, the most normal: taking nightshifts and feeling constantly tired, or taking nightshifts and not feeling constantly tired? or are those equally common and normal?

Participants who chose the first, second, and third answer were assumed to endorse either one of the following statements: $Nightshifts \mid\!\sim Tired$; $Nightshifts \mid\!\sim \neg Tired$; and $(Nightshifts \not\mid\!\sim Tired) \wedge (Nightshifts \not\mid\!\sim \neg Tired)$, respectively.

The fourth, fifth, and sixth questions measured participants' ascriptions of causality and facilitation between (i) taking nightshifts and feeling constantly tired, (ii) working under a stressful boss and feeling constantly tired, and (iii) taking nightshifts under a stressful boss and feeling constantly tired. For example, the fourth question read:

> Fill in the blank with the word 'caused' or 'facilitated', as seems the most appropriate. If neither seems appropriate, fill in the blank with 'xxx': *Taking nightshifts ... the fact that Benoît feels constantly tired.*

The whole process was then repeated with a second and a third story. In the second story, the character took nightshifts and had become a dad. In the third story, the character had a stressful boss and had become a dad. Thus, overall, the three characters were described as constantly feeling very tired (an uncommon feeling for them) after two recent changes in their lives, taken from a pool of three: taking nightshifts, having a stressful boss, becoming a dad. For each character, the first three questions assessed participants' background knowledge with respect to (i) the relation between the first event and feeling constantly tired; (ii) the second event and feeling constantly tired; and (iii) the conjunction of the two events and feeling constantly tired. Then, the fourth, fifth, and sixth questions assessed participants ascriptions of causality or facilitation between (i) the first event and feeling constantly tired; (ii) the second event and feeling constantly tired; and (iii) the conjunction of the two events and feeling constantly tired.

The experiment was conducted in French,[4] and the order in which the stories were presented to the participants was counterbalanced (i.e., they were presented in one order to half of the participants and in the opposite order to the other half).

#### 4.2.2. Results

Out of the 116 ascriptions that the model predicted to be of facilitation, 68% indeed were so, 11% were of causality, and 21% were neither. Out of the 224 ascriptions that the model predicted to be of causality, 46% indeed were, 52% were of facilitation, and 2% were neither. (Remember that what is meant by 'ascription' is a choice of term – the ascription is of causality for participants who selected the term 'caused', it is of facilitation for participants who selected the term 'facilitated' and the ascription is blank for participants who declined to choose a term.) The global trend in the results is thus that background knowledge that theoretically matches a facilitation ascription indeed largely leads people to make such an ascription, while background knowledge that theoretically matches a causality ascription leads people to divide equally between causality and facilitation ascriptions. This trend is statistically reliable for almost all ascriptions required by the task. Relevant statistics ($\chi^2$ scores) are higher than 7.7 for seven out of the nine ascriptions ($p<.05$, one-tailed, in all cases), and higher than 3.2 for the remaining two ascriptions ($p<.10$, one-tailed, in both cases). From these results, it appears that the notion of facilitation does have intuitive appeal to lay reasoners, and that it is broadly used as defined in our model. In particular, it clearly has a role to play in situations where an ascription of causality sounds too strong a conclusion, but no ascription at all sounds not strong enough.

### 4.3. Experiment 2

Experiment 2 was designed to consolidate the results of Experiment 1 and to answer the following questions: Does the fact that background knowledge matches Definition 1 or Definition 12 affect the strength of the link participants perceive between two reported events, and does this perceived strength in turn determine whether they make an ascription of causality or facilitation?

[4] The phrase 'a favorisé' was used for 'facilitated', instead of the apparently straightforward translation 'a facilité', for it seemed pragmatically awkward to use the French verb 'faciliter' for an undesirable outcome like being constantly tired.

Author's personal copy



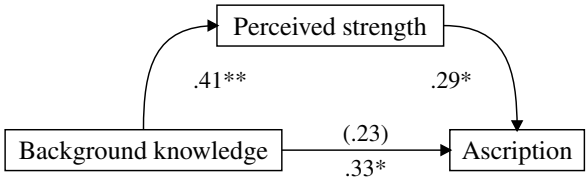


Fig. 1. Mediating role of perceived strength for the effect of background knowledge on ascription. Coefficients are standardized $\beta$s, $^{*}p < .05$, $^{**}p < .01$.

#### 4.3.1. Methods

Participants were 41 undergraduates. Elements of their background knowledge were assessed as in Experiment 1, in order to select triples of propositions $\langle Context, Factor, Effect\rangle$ that matched either Definition 1 or Definition 12. For example, a participant might believe that one has generally no accident when driving, but that one will generally have an accident when driving after some serious drinking; for this participant, $\langle Drive, SeriousDrinking, Accident\rangle$ is a match with Definition 1. Experiment 2 used a richer variety of themes than Experiment 1, including alcohol and road accidents, the appropriate preparation of tea, and the way smoking can deteriorate one's sensitivity to subtle flavors.

Once background beliefs regarding each triple were assessed, participants rated on a 9-point scale how strongly *Factor* and *Effect* were related. Finally, as a measure of ascription, they chose an appropriate term to describe the relation between *Factor* and *Effect*, from a list including *causes*, *facilitates*, *refutes*, *explains*, *justifies*, and *is independent of*.

#### 4.3.2. Results

Out of the 16 ascriptions that the model predicted to be of facilitation, 14 were so, and two were of causality. Out of the 25 ascriptions that the model predicted to be of causality, 11 were so, and 14 were of facilitation. Beliefs thus had the expected influence on ascriptions, $\chi^2 = 4.5$, $p < .05$. The trend observed in Experiment 1 is replicated in Experiment 2. We also conducted a *mediation analysis* of our data, which consists in a series of three regression analyzes (see Fig. 1).

In all regression analyzes, background knowledge was encoded as $-1$ when it matched Definition 12, and as $+1$ when it matched Definition 1. Ascriptions of facilitation were encoded as $-1$, and ascriptions of causality were encoded as $+1$. The first regression assessed the effect of background knowledge on ascription, which was statistically significant, $\beta = .33$, $p < .05$. The second regression assessed the effect of background knowledge on perceived strength, which was also significant, $\beta = .41$, $p<.01$. In the third regression, background knowledge and perceived strength were entered simultaneously. Perceived strength was a reliable predictor of ascription, $\beta = .29$, $p<.05$, which was no longer the case for background knowledge, $\beta = .23$, $p > .05$. Data thus meet the requirements of a mediational effect. We can therefore conclude that whether the background knowledge of participants matches Definition 12 or Definition 1 determines their final ascription of $C : Factor \Rightarrow_{\mathrm{fa}} Effect$ or $C : Factor \Rightarrow_{\mathrm{ca}} Effect$ through its effect on the perceived strength of the link between *Factor* and *Effect*.

The two experiments we have reported show that our basic definitions of causality and facilitation ascriptions have some degree of descriptive validity. Human subjects do differentiate between causality and facilitation, and broadly along the lines featured in our definitions. The next section will introduce a third notion besides cause and facilitation, namely, that of justification – but it will first discuss our model in relation to previous work on causality.

## 5. Related works

In this section, we first compare our approach to causality with previous works not focused on diagnosis. Then we show a distinction between causality ascription and the notion of explanation.

Author's personal copy



### 5.1. Causality, non-monotonicity, intervention

One of the earliest formal account of causality is due to Von Wright [36]. According to this author, an action caused $p$ to be true if and only if either:

- $p$ was false before the action, and had the action not been taken, $p$ would not have become true, or
- the action maintains $p$ true against the normal course of things, thus preventing $p$ from becoming false.

The first situation straightforwardly relates to Definition 1. The second situation can also be represented in our setting: $B_t$ is known to be true, and after $A_t$ takes place $B_{t+k}$ is still true, although in the normal course of things, had $A$ not happened, $B$ would have become false, i.e., $B_t \wedge C \mid\!\sim \neg B_{t+k}$. The agent knowledge also includes $B_t \wedge A_t \wedge C \mid\!\sim B_{t+k}$. Letting $C' = B_t \wedge C$, this can be rewritten $C' \mid\!\sim \neg B_{t+k}$ and $A \wedge C' \mid\!\sim B_{t+k}$, which is formally Definition 1.

Our approach is in fact directly inspired by previous works on epistemic qualitative independence in the setting of possibility theory, especially [12]. In this paper, believing $B$ is said to be independent of $A$ when learning $A$ does not affect our belief in $B$. At the opposite, $A$ is said to refute $B$ when learning $A$ does turns our belief in $B$ into believing $\neg B$. Our notion of causality is directly based on the latter notion, in the scope of analyzing a sequence of reported events, and using system P instead of possibility theory. When $B$ is believed and learning $A$ only leads to drop this belief, $A$ is said to 'cancel' $B$ in [12]. This notion is instrumental in our definition of facilitation, in a setting that is mathematically equally expressive to possibility theory.

Where our qualitative approach represents the knowledge underlying causal ascriptions by means of non-monotonic consequence relationships, quantitative approaches would represent knowledge by means of structural equations. Following [5], Halpern and Pearl [37,38] have proposed a model that distinguishes real causes (cause in fact) from potential causes, by using an a priori distinction between 'endogenous' variables (the possible values of which are governed by structural equations, for example physical laws), and 'exogenous' variables (determined by external factors). Exogenous variables cannot be deemed causal. Halpern and Pearl's definition of causality formalizes the notion of an active causal process. More precisely, the fact $A$ that a subset of endogenous variables has taken some definite values is the real cause of an event $B$ if (i) $A$ and $B$ are true in the real world, (ii) this subset is minimal, (iii) another value assignment to this subset would make $B$ false, the values of the other endogenous variables that do not directly participate to the occurrence of $B$ being fixed in some manner and (iv) $A$ alone is enough for $B$ to occur in this context. This approach, thanks to the richness of background knowledge when it is represented in structural equations, makes it possible to treat especially difficult examples.[5]

Our model is however not to be construed as an alternative or a competitor to models based on structural equations, like the one of Halpern and Pearl. Indeed, we see our approach as a complement to structural equation modeling. One might not have access to the accurate information needed to build a structural equation model; in this case, our less demanding model might still be operable. Alternatively, a decision-support system may be able to build a structural equation model of the situation, although its users only have access to qualitative knowledge. In that case, the system will be able to compare its own causality ascriptions to the conclusions of the qualitative model, and take appropriate explanatory steps, would those ascriptions be too different. Indeed, our model does not aim at identifying the true, objective cause of an event, but rather at predicting what causal ascription an agent would make based on the limited information it has at its disposal.

Models based on structural equations are often supplemented with the useful notion of *intervention* [5]. In many situations, finding the cause of an event will be much easier if the agent can directly intervene in the manner of an experimenter. In future work, we intend to explore the possibility of supplementing our own model with a similar notion by means of a **do**(•) operator. As for now, we only give a brief example that

[5] Ascribing causality when analyzing a set of reported facts may find its motivation in the search for responsibility behind the occurrence of these facts. Building upon the notion of potential cause, Chockler and Halpern [39] have introduced definitions of responsibility and blame: The extent to which a cause (or an agent) is responsible for an effect is graded, and depends on the presence of other potential causes (or agents). Clearly, the assessment of responsibility from identification of causal relationships raises further problems that are beyond the scope of this article.

Author's personal copy



suggests that our approach needs the complementary notion of intervention for cleaner ascriptions of causality. A stricter condition for an ascription of causality $C$: $A \Rightarrow_{ca} B$ (respectively, facilitation) would be to require that the background knowledge part of Definition 1 (respectively, Definition 12) applies to $\mathbf{do}(A)$, $B$, $C$, where $\mathbf{do}(A)$ means that the occurrence of $A$ is forced by an intervention – thus requiring that the definitions take into account the three distinct components that are: factual observations, pieces of beliefs, and known interventions.

**Example 16** (*Yellow teeth*). An agent learns that someone took up smoking, that this person's teeth yellowed, and that this person developed lung cancer. The agent believes that generally speaking, it is abnormal to be a smoker, to have yellow teeth, and to develop lung cancer (respectively, $C \mid\!\sim \neg Smoke$, $C \mid\!\sim \neg Yellow$, $C \mid\!\sim \neg Lung$). The agent believes that it is normal for smokers to have yellow teeth ($C \wedge Smoke \mid\!\sim Yellow$) and to develop lung cancer ($C \wedge Smoke \mid\!\sim Lung$), and that it is not abnormal for someone who has yellow teeth to develop lung cancer ($C \wedge Yellow \mid\!\sim \neg Lung$). From these beliefs and observations, Definitions 12 and 1 would allow for various ascriptions, including the following one: Smoking caused the yellow teeth which *in turn* facilitated lung cancer. With the additional constraint based on the $\mathbf{do}(\bullet)$ operator, only one set of ascriptions remains possible: *Both* the yellow teeth and the lung cancer were caused by smoking. Yellow teeth cannot be said anymore to facilitate lung cancer because, inasmuch as lung cancer is generally abnormal, it holds that $C \wedge \mathbf{do}(Yellow) \mid\!\sim \neg\ Lung$: There is no reason to think that one will develop lung cancer after painting one's teeth yellow.

### 5.2. Causality vs. justification

Perceived causality as expressed in Definition 1 should be distinguished from the following situation, which we term 'justification.' We write that $C : A \Rightarrow_{ju} B$ when an agent judges that the occurrence of $A$ in context $C$ gave reason to expect the occurrence of $B$.

**Definition 17** (*Justification*). Let us assume that an agent learns of the sequence $\neg B_t$, $A_t$, $B_{t+k}$. Let us call $C$ (the *context*) the conjunction of all other facts known by or reported to the agent at time $t$. If the agent possesses the piece of default knowledge that $A \wedge C \mid\!\sim B$, and if it holds for the agent that $C \not\mid\!\sim \neg B$ and $C \not\mid\!\sim B$, the agent will perceive $A_t$ to justify the expectation that $B_{t+k}$ would occur in context $C_t$, denoted $C_t : A_t \Rightarrow_{ju} B_{t+k}$.

What we call justification is borrowed again from [12] and is akin to the notion of explanation following Spohn [40]: Namely, '$A$ is a reason for $B$' when raising the epistemic rank for $A$ raises the epistemic rank for $B$. Gärdenfors [41] captured this view to some extent, assuming that $A$ is a reason for $B$ if $B$ is not retained in the contraction of $A$. Williams et al. [42] could account for the Spohnian view in a more refined way using kappa-rankings and transmutations, distinguishing between weak and strong explanations. As our framework can easily be given a possibilistic semantics [30], it could properly account for this line of thought, although our distinction between perceived causation and epistemic justification is not the topic of the above works.

In our model this distinction is very clear. Faced with facts $C$, $\neg B_t$, $A_t$, $B_{t+k}$, an agent believing that $C \not\mid\!\sim \neg B$, $C \not\mid\!\sim B$ and $A \wedge C \mid\!\sim B$ may doubt that the change from $\neg B_t$ to $B_{t+k}$ is really due to $A_t$, although the latter is indeed the very reason for the lack of surprise at having $B_{t+k}$ reported. Indeed, situation $\neg B_t$ at time $t$ appears to the agent to be contingent, since it is neither a normal nor an abnormal course of things in context $C$. This clearly departs from the situation where $C \mid\!\sim \neg B$ and $A \wedge C \mid\!\sim B$, wherein the agent will judge that $C : A \Rightarrow_{ca} B$. In a nutshell, the case whereby $C \not\mid\!\sim \neg B$, $C \not\mid\!\sim B$ and $A \wedge C \mid\!\sim B$ cannot be interpreted as the recognition of a causal phenomenon by an agent: All that can be said is that reporting $A$ caused the agent to start believing $B$, and that it should not be surprised of having $B_{t+k}$ reported.

## 6. Concluding remarks

We have presented a simple qualitative model of the *causal ascriptions* an agent will make from its background default knowledge, when faced with a series of events. The model assumes that the agent's beliefs are represented in the setting of non-monotonic reasoning, and more precisely in System P. A new notion, less

Author's personal copy



committing than perceived causality, has also been laid bare. It naturally appears in the formal model, provided that the Rational Monotony axiom is added – furthermore, this notion is proved to be cognitively relevant via a set of experimental tests.

The model we have presented is certainly just a first step towards a fully satisfactory account of causal ascription. In a provocatively titled paper (Causality is undefinable), Zadeh [43] illustrated the difficulty of causal ascriptions by means of the following example:

**Example 18** (*From Zadeh* [43]). I am called by a friend. He needs my help and asks me to rush to his home. I jump into my car and drive as fast as I can. At an intersection, I am hit by another car. I am killed. Who caused my death? My friend; I; or the driver of the car that hits me. Note that in such a scenario, it seems possible to expand the list of candidate causes very easily, in an almost endless manner, as here, e.g., "my emotionality that limits my capacities to avoid accidents," "in my hurry, I had not fasten my seat belt," or even "the fact that the phone was working and I was there for receiving the call."

While we do not claim that our model can entirely take care of such an example, we note that it might well handle some of its crucial aspects. For instance, not fastening the seat belt would likely count as a facilitation, rather than as a cause of death. Picking up the phone would likely not be considered a cause of death, because of the restricted transitivity of causal ascriptions in our model. Suppose that picking up the phone caused listening to the friend's story, which caused the fast driving, which caused the accident, which caused the death. Did picking up the phone caused the death? In our model, for such a conclusion to hold, one would need to accept that traffic accidents are usually diagnostic of fast driving, that fast driving is usually diagnostic of having listened to a friend's call for help, and that having listened to a friend's call for help is usually diagnostic of having picked up the phone. At least the second link in that chain is very doubtful.

Future developments of our model should include the formal properties of facilitation in the formal approach, and study the potential of the notion of intervention in the model. In addition to supplementing this model with a **do**(•) operator, we intend to extend our present work in three main directions. First, we should be able to equip our framework with possibilistic qualitative counterparts to Bayesian networks [44], since System P augmented with Rational Monotony can be represented in possibilistic logic [30].[6] Second, we should be able to derive postulates for causality and facilitation from the independence postulates presented in [12]. Finally, in parallel to further theoretical elaboration, we will maintain a systematic experimental program that will test the psychological plausibility of our definitions, properties, and postulates.

## References


[1] Y. Peng, J.A. Reggia, Abductive Inference Models for Diagnostic Problem-Solving, Springer Verlag, Berlin, 1990.
[2] R. Reiter, A theory of diagnosis from first principles, Artificial Intelligence 32 (1987) 57–95.
[3] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, CA, 1988.
[4] D. Dubois, H. Prade, Probability theory in artificial intelligence. Book review of J. Pearl's 'Probabilistic Reasoning in Intelligent Systems', Journal of Mathematical Psychology 34 (1999) 472–482.
[5] J. Pearl, Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, 2000.
[6] J. de Kleer, J.S. Brown, Theories of causal ordering, Artificial Intelligence 29 (1986) 33–61.
[7] E. Giunchiglia, J. Lee, N. McCain, V. Lifschitz, H. Turner, Non-monotonic causal theories, Artificial Intelligence 153 (2004) 49–104.
[8] N. McCain, H. Turner, A causal theory of ramifications and qualifications, in: C.S. Mellish (Ed.), Proceedings of the 14th International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal, Québec, Canada, August 20–25, Morgan Kaufmann, San Francisco, CA, 1995, pp. 1978–1984.
[9] H. Turner, A logic of universal causation, Artificial Intelligence 113 (1999) 87–123.
[10] H. Geffner, Causal theories for nonmonotonic reasoning, in: Proceedings of the 8th National Conference on Artificial Intelligence. Boston, Massachusetts, July 29–August 3, AAAI Press, Boston, MA, 1990, pp. 524–530.
[11] G. Shafer, Causal logic, in: H. Prade (Ed.), Proceedings of the 13th European Conference on Artificial Intelligence, Brighton, UK, August 23–28, Wiley, Chichester, England, 1998, pp. 711–719.
[12] D. Dubois, L. Fariñas Del Cerro, A. Herzig, H. Prade, A roadmap of qualitative independence, in: D. Dubois, H. Prade, E.P. Klement (Eds.), Fuzzy Sets Logics and Reasoning about Knowledge, Applied Logic series, vol. 15, Kluwer, Dordrecht, The Netherlands, 1999, pp. 325–350.


[6] This raises the more general question of the possibility of reading causality and facilitation ascriptions (in the sense used in this article) from a Bayesian net structure, or building such a probabilistic or possibilistic net from such ascriptions.

Author's personal copy



[13] D. Dubois, H. Prade, Causality and nonmonotonicity, in: Proceedings of the International Conference on Advances in Intelligent Systems – Theory and Applications (AISTA'04) , Luxembourg, November 11–18, 2004.
[14] D. Dubois, H. Prade, Modeling the role of (ab)normality in the ascription of causality judgements by agents, in: L. Morgenstern, M. Pagnucco (Eds.), Proceedings of IJCAI-05 Workshop on Nonmonotonic Reasoning, Action, and Change (NRAC'05), Edinburg, Scotland, August 1, 2005, pp. 22–27.
[15] J.F. Bonnefon, R.M. Da Silva Neves, D. Dubois, H. Prade, Background default knowledge and causality ascriptions, in: G. Brewka, S. Coradeschi, A. Perini, P. Traverso (Eds.), Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06), Riva del Garda, Italy, August 29–September 1, IOS Press, Zurich, 2006, pp. 11–15.
[16] J.F. Bonnefon, R.M. Da Silva Neves, D. Dubois, H. Prade, Model and experimental studies of causality ascriptions, in: J. Dix, A. Hunter (Eds.), Proceedings of the 11th Workshop on Nonmonotonic Reasoning (NMR06), Lake District, UK, May 30–June 1.
[17] D. Kayser, F. Nouioua, Representing knowledge about norms, in: R. López de Mántaras, L. Saitta (Eds.), Proceedings of the 16th European Conference on Artificial Intelligence. Valencia, Spain, August 22–27, IOS Press, Zurich, 2004, pp. 363–367.
[18] D. Kayser, F. Nouioua, About norms and causes, International Journal of Artificial Intelligence Tools 14 (2005) 7–24.
[19] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1990) 167–207.
[20] D. Lehmann, M. Magidor, What does a conditional knowledge base entail? Artificial Intelligence 55 (1992) 1–60.
[21] S. Benferhat, J.F. Bonnefon, R.M. Da Silva Neves, An experimental analysis of possibilistic default reasoning, in: D. Dubois, C.A. Welty, M.-A. Williams (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference (KR2004), Whistler, Canada, June 2–5, AAAI Press, 2004, pp. 130–140.
[22] S. Benferhat, J.F. Bonnefon, R.M. Da Silva Neves, An overview of possibilistic handling of default reasoning, with experimental studies, Synthese 146 (2005) 53–70.
[23] R.M. Da Silva Neves, J.F. Bonnefon, E. Raufaste, An empirical test for patterns of nonmonotonic inference, Annals of Mathematics and Artificial Intelligence 34 (2002) 107–130.
[24] M. Ford, System LS: a three tiered nonmonotonic reasoning system, Computational Intelligence 20 (2004) 89–108.
[25] N. Pfeifer, G.D. Kleiter, Coherence and nonmonotonicity in human reasoning, Synthese 146 (2005) 93–109.
[26] Y. Shoham, Nonmonotonic reasoning and causation, Cognitive Science 14 (1990) 213–252.
[27] E. Adams, The logic of conditionals, Inquiry 8 (1965) 166–197.
[28] S. Benferhat, D. Dubois, H. Prade, Possibilistic and standard probabilistic semantics of conditional knowledge bases, Journal of Logic and Computation 9 (1999) 873–895.
[29] P. Snow, Diverse confidence levels in a probabilistic semantics for conditional logics, Artificial Intelligence 113 (1999) 269–279.
[30] S. Benferhat, D. Dubois, H. Prade, Nonmonotonic reasoning, conditional objects and possibility theory, Artificial Intelligence 92 (1997) 259–276.
[31] A. Bochman, A logic for causal reasoning, in: G. Gottlob, T. Walsh (Eds.), Proceedings of the 8th International Joint Conference in Artificial Intelligence (IJCAI 2003), August 9–15, Morgan Kaufmann, Acapulco, Mexico, 2003, pp. 141–146.
[32] D.J. Hilton, B.R. Slugoski, Knowledge-based causal attribution: the abnormal conditions focus model, Psychological Review 93 (1986) 75–88.
[33] I. Gavansky, G.L. Wells, Counterfactual processing of normal and exceptional events, Journal of Experimental Social Psychology 25 (1989) 314–325.
[34] H.L.A. Hart, T. Honoré, Causation in the Law, Oxford University Press, Oxford, 1985.
[35] J. Bell, Causation as production, in: G. Brewka, S. Coradeschi, A. Perini, P. Traverso (Eds.), Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06), August 29–September 1, IOS Press, Rivadel Garda, Italy, Zurich, 2006, pp. 327–331.
[36] G.H. von Wright, Norm and Action: A Logical Enquiry, Routledge, London, 1963.
[37] J. Halpern, J. Pearl, Causes and explanations: a structural-model approach – part 1: Causes, British Journal for the Philosophy of Science 56 (2005) 843–887.
[38] J. Halpern, J. Pearl, Causes and explanations: a structural-model approach – part 2: Explanations, British Journal for the Philosophy of Science 56 (2005) 889–911.
[39] H. Chockler, J. Halpern, Responsibility and blame. A structural-model approach, in: IJCAI'03, Morgan Kaufmann, San Francisco, CA, 2003, pp. 147–153.
[40] W. Spohn, Deterministic and probabilistic reasons and causes, Erkenntnis 19 (1983) 371–393.
[41] P. Gärdenfors, The dynamics of belief systems: foundations vs. coherence theories, Revue Internationale de Philosophie 44 (1990) 24–46.
[42] M.-A. Williams, M. Pagnucco, N. Foo, B. Sims, Determining explanations using transmutations, in: C.S. Mellish (Ed.), Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal, Québec, Canada, August 20–25, Morgan Kaufmann, San Francisco, CA, 1995, pp. 822–830.
[43] L.A. Zadeh, Causality is indefinable – towards a theory of hierarchical definability, in: J.A. Meech, M.M. Veiga, Y. Kawazoe, S.R. LeClair (Eds.), Intelligence in a Materials World: Selected Papers from IPMM-2001, CRC Press, Boca Raton, FL, 2002, pp. 29–34.
[44] S. Benferhat, D. Dubois, L. Garcia, H. Prade, On the transformation between possibilistic logic bases and possibilistic causal networks, International Journal of Approximate Reasoning 29 (2002) 135–173.