RESEARCH-ARTICLE

A Causal Model Checker for Legal Cases

RŪTA LIEPIŅA, University of Bologna, Bologna, BO, Italy

TIAGO DE LIMA, Artois University, Arras, Hauts-de-France, France

EMILIANO LORINI, University of Toulouse, Toulouse, Occitanie, France

GIUSEPPE PISANO, University of Bologna, Bologna, BO, Italy

GIOVANNI SARTOR, University of Bologna, Bologna, BO, Italy

Published: 13 January 2026

ICAIL 2025: 20th International Conference on Artificial Intelligence and Law
June 16 - 20, 2025
IL, Chicago, USA

ICAIL '25: Proceedings of the Twentieth International Conference on Artificial Intelligence and Law (June 2025)
https://doi.org/10.1145/3769126.3769207
ISBN: 9798400719394

A Causal Model Checker for Legal Cases

Rūta Liepiņa
ALMA-AI
University of Bologna
Bologna, Italy
ruta.liepina@unibo.it

Tiago de Lima
CRIL
University of Artois
Lens, France
CRIL
CNRS
Lens, France
delima@cril-lab.fr

Emiliano Lorini
IRIT, CNRS
Toulouse University
Toulouse, France
Emiliano.Lorini@irit.fr

Giuseppe Pisano
ALMA-AI
University of Bologna
Bologna, Italy
g.pisano@unibo.it

Giovanni Sartor
ALMA-AI
University of Bologna
Bologna, Italy
Law Department
European University Institute
Florence, Italy
giovanni.sartor@unibo.it

Abstract

Causation plays a central role in the attribution of responsibility, especially in the legal domain, where complex causal scenarios frequently arise. Traditionally, legal reasoners have relied on the idea that a cause must be a necessary condition of its effect, which falls short in scenarios involving overdetermination, preemption, or omission, thereby failing to adequately identify causes-in-fact. In this paper, we present a novel analysis of selected legal cases, each exemplifying common causal dilemmas discussed in causal literature. We employ three different notions of cause in our analysis: abductive explanation (AXp), the NESS test (Necessary Element of a Sufficient Set) and actual cause. We express the three notions and some of their variants in a modal language for causal reasoning that we interpret on a rule-based semantics. We provide a model checking algorithm for our modal language relying on a reduction into TQBF as well as an implementation of the legal cases in our causal model checker to automatically verify "what is the cause of what" and what types of causes apply in each legal case. Our interdisciplinary approach highlights the usefulness of logic-based methods for legal analysis, offering a fully transparent model-checking toolbox that could potentially support legal reasoners in disentangling complex factual scenarios.

CCS Concepts

• **Theory of computation → Modal and temporal logics**; **Automated reasoning**; • **Applied computing → Law**.

ICAIL 2025, Chicago, IL, USA

ACM ISBN 979-8-4007-1939-4/25/06
https://doi.org/10.1145/3769126.3769207

Keywords

Knowledge representation and reasoning, Base-based modal logic, Causation, Model checking,

ACM Reference Format:
Rūta Liepiņa, Tiago de Lima, Emiliano Lorini, Giuseppe Pisano, and Giovanni Sartor. 2025. A Causal Model Checker for Legal Cases. In *20th International Conference on Artificial Intelligence and Law (ICAIL 2025), June 16–20, 2025, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3769126.3769207

1 Introduction

Causation plays a central role in the attribution of responsibility, especially in the legal domain. However, the concept of causality in law remains highly debated, especially in complex causal scenarios. The analysis of causality in the legal domain is often affected by too restrictive and insular approaches. On the one hand, legal reasoners often rely on the traditional view that the cause of an effect must be a necessary condition (*conditio since qua non*) of it, an approach which cannot handle cases of overdetermination and preemption. On the other hand, notions of causality have been proposed that also include some usual requirements for legal liability (such as the foreseability, non-exceptionality, or increased risk of the effect, given the cause [8]). Following the approach by R.W. Wright [18], we believe that more clarity and interdisciplinary insight can be obtained by relying upon a general notion of causation, which corresponds to natural causation, to which we refer as "cause in fact." A more transparent analysis of causation in the law is crucial not only for enhancing legal certainty and coherence but also for deepening our understanding of legal reasoning in new cases.

In this paper, we present a novel analysis of six selected legal cases, exemplifying the following controversial aspects pertaining to causation in law (but also to other domains): i) **overdetermination**, where different causal paths independently lead to the same effect; ii) **preemption**, where two causal paths interfere, one preventing the other from achieving the effect; iii) **confounding**,

where a common cause determines two different effects, generating a correlation between them; iv) **omission**, where the effect is caused by the non-occurrence of an event; and v) **ennoblement**, where both the occurrence and non-occurrence of an event cause an effect through different paths.

We employ three different notions of cause: abductive explanation (AXp), NESS (Necessary Element of a Sufficient Set), and actual cause. We express these notions and some of their variants in a modal language for causal reasoning that we interpret on a rule-based semantics in which causal information is expressed in equational form. We provide a model checking algorithm relying on a reduction into TQBF (True QBF) [2] as well as an implementation of the legal cases in our model checker to automatically verify "what is the cause of what" and what notions of a cause apply in each legal case. Automatic verification through model checking supports and certifies our analysis of how each notion handles the previous five problematic aspects of causation. Our interdisciplinary approach highlights the usefulness of logic-based methods for legal analysis, offering a fully transparent model checking toolbox that could potentially support legal practitioners and judges in disentangling complex factual scenarios. Technical proofs are given in the appendix. The source code of the causal model checker for legal cases is available online.[1]

The paper is organised as follows. We begin with a running example of preemption – a criminal case of attempted murder in which the offender attempted to kill an already deceased victim. In Section 3, we present the logical framework and the model checking approach based on it. Then in Section 4, we show how to express the different notions of cause in the language. In Sections 5 and 6, we analyse causation in legal cases and compare various causal approaches, discussing their strengths and weaknesses in addressing causal complexities.

2 Running Example

As a running example, we will use the *Dlugash v. People*, to which we will refer to illustrate the formal definitions[2]:

> In 1973, the victim, Mr. Geller, was found shot to death in his Brooklyn apartment. Bush had shot first and fatally wounded the victim, whereas Dlugash had fired several shots after Geller had already died. Bush was considered to have caused the death and consequently convicted for murder, Dlugash was only convicted for attempted murder, since he tried to kill Geller, with suitable means, but failed to do so.

We use the following abbreviations for the relevant propositions representing the atomic facts.

- *Bd*/*Dd*: Bush/Dlugash decides to shoot Geller (exogenous).
- *Bs*/*Ds*: Bush/Dlugash shoots Geller.
- *Bk*/*Dk*: Bush/Dlugash kills Geller.
- *Di*: Geller dies.

This case is especially interesting since it concerns preemption: Dlugash could not kill Geller by shooting his body since Geller had already been killed by Bush. This is why Bush was convicted for murder while Dlugash only for attempted murder. Preemption is interesting since it challenges most traditional approaches to causation [5]: both putative causes would be sufficient for the event, in the absence of the other, and neither is necessary for it.

[1] https://gitlab.in2p3.fr/tiago.delima/causal-model-checker

[2] *Dlugash v. People of State of NY*, 476 F. Supp. 921 (E.D.N.Y. 1979).

3 Formal Preliminaries

In this section, we introduce a fragment of the modal language for causal reasoning proposed in [14]. It includes i) a causal necessity modality to represent causally necessary facts, and ii) an interventionist modality to represent causally necessary facts after an intervention has taken place (i.e., causal necessity *post intervention*). The modal language was recently extended in [4] with general modalities of *conditional* causal necessity (i.e., causal necessity conditional on the occurrence of a causal change operation). After having introduced a rule-based semantics in which causal information is represented in an equational form, we show how to interpret the modal language over it. We conclude the section by recalling the model checking problem for our language. We provide a polysize reduction of it into TQBF which is an adaptation of the polysize reduction for the language of conditional causal necessity given in [4].

3.1 Semantics

Let $\mathbb{P}$ be a countable set of atomic propositions whose elements are noted $p, q, \ldots$ We note $\mathcal{L}_{\text{PROP}}$ the propositional language built from $\mathbb{P}$. In particular, $\mathcal{L}_{\text{PROP}}$ includes the atomic propositions in $\mathbb{P}$ and the formulas built from the two usual Boolean connectives $\neg$ (negation) and $\wedge$ (conjunction) from which the other Boolean connectives can be constructed including $\vee$ (disjunction), $\rightarrow$ (implication), $\leftrightarrow$ (double implication), $\top$ (tautology) and $\bot$ (contradiction). Elements of $\mathcal{L}_{\text{PROP}}$ are noted $\omega, \omega', \ldots$ Given $\omega \in \mathcal{L}_{\text{PROP}}$, we note with $\mathbb{P}(\omega)$ the set of atomic propositions occurring in ω. Moreover, if $X \subseteq \mathcal{L}_{\text{PROP}}$ then $\mathbb{P}(X) = \bigcup_{\omega \in X} \mathbb{P}(\omega)$.

Formulas of the propositional language $\mathcal{L}_{\text{PROP}}$ are interpreted in the usual relative to a propositional valuation $V \subseteq \mathbb{P}$, as follows:

$$\begin{array}{lll} V \models p & \text{iff} & p \in V, \\ V \models \neg\omega & \text{iff} & V \not\models \omega, \\ V \models \omega_1 \wedge \omega_2 & \text{iff} & V \models \omega_1 \text{ and } V \models \omega_2. \end{array}$$

In order to represent causal information, we consider propositional formulas in equational form. An equational formula for a proposition p is a propositional formula of the form $p \leftrightarrow \omega$ which unambiguously specifies the truth value of p using a propositional formula ω made of propositions other than p. We note $\mathcal{L}_{\text{EQ}}$ the corresponding set of equational formulas:

$$\mathcal{L}_{\text{EQ}} = \Big\{ p \leftrightarrow \omega : p \in \mathbb{P}, \omega \in \mathcal{L}_{\text{PROP}};\ \text{and } p \notin \mathbb{P}(\omega) \Big\}.$$

For every $p \in \mathbb{P}$, $\mathcal{L}_{\text{EQ}}(p)$ is the set of equational formulas for p. For notational convenience, elements of $\mathcal{L}_{\text{EQ}}$ are also noted $\epsilon, \epsilon', \ldots$

The main constituent of our semantics is the following notion of equational state: a finite set of equational formulas, called causal base, supplemented with a propositional valuation that is compatible with it. For the sake of exposition, here we only consider causal bases containing information in equational form. The semantics given in [14] is more general as it allows causal bases to include arbitrary propositional formulas.

Definition 1 (Equational state). *An equational state is a pair $S = (C, V)$, with $C \subseteq \mathcal{L}_{\text{EQ}}$ finite, $V \subseteq \mathbb{P}(C)$ and such that*

i) $\forall p \leftrightarrow \omega \in C, V \models p \leftrightarrow \omega$,
ii) $\forall p \in \mathbb{P}, \forall p \leftrightarrow \omega, p \leftrightarrow \omega' \in C, \omega = \omega'$.

The set of equational states is noted $\mathbf{S}_{Eq}$.

The propositional valuation V represents the actual environment (or situation), while C represents the base of causal information (viz. the causal base). According to the condition i), V must be compatible with C, that is, if the causal information $p \leftrightarrow \omega$ is included in the actual causal base (i.e., $p \leftrightarrow \omega \in C$) then it should be true in the actual environment (i.e., $V \models p \leftrightarrow \omega$). According to the condition ii), a causal base should contain at most one equational formula for each atomic proposition.

Let us illustrate the components of our formal semantics with the help of the running example we introduced in Section 2.

Example 1. *The People v. Dlugash case can be represented by an equational state $S_0 = (C_0, V_0)$ such that:*
$C_0 = \{Di \leftrightarrow Bk \vee Dk, Bk \leftrightarrow Bs, Dk \leftrightarrow (Ds \wedge \neg Bk),$
$Bs \leftrightarrow Bd, Ds \leftrightarrow Dd\}$
$V_0 = \{Di, Bk, Bs, Ds, Bd, Dd\}$.
The causal rule $Di \leftrightarrow Bk \vee Dk$ expresses the fact that Geller dies if and only if he is killed by Bush or Dlugash, the rule $Bk \leftrightarrow Bs$ expresses the fact that Geller is killed by Bush if and only if Bush shoots him, the rule $Dk \leftrightarrow (Ds \wedge \neg Bk)$ expresses the fact that Geller is killed by Dlugash if and only if Dlugash shoots him but Bush does not kill him, and so on. According to the propositional valuation V_0, in the actual situation Bush decides to shoot Geller, Bush shoots Geller, Bush kills Geller, Geller dies, Duglash decides to shoot, Duglash shoots Geller, but Duglash does not kill Geller. Notice that, in agreement with Definition 1, for every equational rule in the causal base C_0 the propositional valuation V_0 is compatible with it. That is, we have $V_0 \models Di \leftrightarrow Bk \vee Dk$, $V_0 \models Bk \leftrightarrow Bs$, $V_0 \models Dk \leftrightarrow (Ds \wedge \neg Bk)$, $V_0 \models Bs \leftrightarrow Bd$ and $V_0 \models Ds \leftrightarrow Dd$.

From an equational state, it is straightforward to extract a set of endogenous variables and a set of exogenous ones. A variable is endogenous if there is an equational formula for it in the actual causal base, it is exogenous if it appears in the actual causal base but there is no equational formula for it. In formal terms,

Definition 2 (Exogenous and endogenous variables). *Let $S = (C, V)$ be an equational state. Its set of exogenous variables $exo(S)$ and its set of endogenous variables $end(S)$ are defined as follows:*

$$end(S) = \big\{p \in \mathbb{P}(C) : \exists \omega \in \mathcal{L}_{\text{PROP}}(\mathbb{P} \setminus \{p\}) \text{ s.t. } p \leftrightarrow \omega \in C\big\},$$
$$exo(S) = \mathbb{P}(C) \setminus end(S).$$

The following definition introduces the notion of causal compatibility.

Definition 3 (Causal compatibility). *We define $\equiv$ to be the binary relation on the set $\mathbf{S}_{Eq}$ such that, for every $S = (C, V), S' = (C', V') \in \mathbf{S}_{Eq}$:*

$$S \equiv S' \text{ if and only if } C = C'.$$

$S \equiv S'$ means that state S and state S' are causally compatible since they share the same causal information.

We conclude this section by defining the concept of intervention. We conceive an intervention as a possibly empty finite set of equational formulas of type $p \leftrightarrow \top$ or $p \leftrightarrow \bot$ with at most one equational formula for each variable. We define the set of interventions as follows:

$$Int = \big\{\{p_1 \leftrightarrow \tau_1, \ldots, p_k \leftrightarrow \tau_k\} : \forall 1 \leq k', k'' \leq k, \text{ if } k' \neq k'' \text{ then } p_{k'} \neq p_{k''} \text{ and } \tau_1, \ldots, \tau_k \in \{\top, \bot\}\big\}.$$

Elements of Int are noted $E, E', \ldots$ For the sake of generality, we sometimes use the generic term event to refer to an intervention. Given $E \in Int$, we define the corresponding conjunction over variables:

$$\widehat{E} = \bigwedge_{p \leftrightarrow \top \in E} p \wedge \bigwedge_{p \leftrightarrow \bot \in E} \neg p.$$

For every finite set of atomic propositions $Z \subseteq \mathbb{P}$, we note Int_Z the set of interventions for Z, that is,

$$Int_Z = \big\{E \in Int : (\forall p \in Z, p \leftrightarrow \top \in E \text{ or } p \leftrightarrow \bot \in E) \text{ and } (\forall p \notin Z, p \leftrightarrow \top \notin E \text{ and } p \leftrightarrow \bot \notin E)\big\}.$$

From a semantic point of view, an intervention $\{p_1 \leftrightarrow \tau_1, \ldots, p_k \leftrightarrow \tau_k\}$ replaces any equational formula for $p_{k'}$ with $1 \leq k' \leq k$ in a causal base by the equational formula $p_{k'} \leftrightarrow \tau_{k'}$. Following this idea, the following definition introduces the notion of causal compatibility post intervention.

Definition 4 (Causal compatibility post intervention). *Let $E \in Int$. We define $\Rightarrow^E$ to be the binary relation on the set of equational states $\mathbf{S}_{Eq}$ such that, for every $S = (C, V), S' = (C', V') \in \mathbf{S}_{Eq}$:*

$$S \Rightarrow^E S' \text{ if and only if } C' = \big(C \setminus \bigcup_{p \in \mathbb{P}(E)} \mathcal{L}_{\text{EQ}}(p)\big) \cup E.$$

$S \Rightarrow^E S'$ means that state $S' = (C', V')$ is compatible with state $S = (C, V)$ after the occurrence of event E. Specifically, the latter is the case if the causal base C' is the result of the following replacement operation applied to the causal base C: first of all remove from C all equational formulas for the propositions on which we intervene through E, and then add to the resulting causal base all equational formulas corresponding to the current complex intervention E. Note that $S \Rightarrow^{\emptyset} S'$ if and only if $S \equiv S'$.

Let us illustrate Definition 4 with the help of Dlugash v. People case whose formalization was given in Example 1 above.

Example 2. *A state $S' = (C', V')$ is compatible with state S_0 after a negative intervention on proposition Bs (e.g., someone unloads Bush's gun), denoted by $S_0 \Rightarrow^{\{Bs \leftrightarrow \bot\}} S'$, if and only if*

$$C' = \big\{Di \leftrightarrow Bk \vee Dk, Bk \leftrightarrow Bs, Dk \leftrightarrow (Ds \wedge \neg Bk), Bs \leftrightarrow \bot, Ds \leftrightarrow Dd\big\}.$$

3.2 Modal Language

We consider the following two-layer modal language for causal reasoning:

$$\mathcal{L}_0 \quad \stackrel{def}{=} \quad \alpha ::= p \mid \top \mid \neg\alpha \mid \alpha \wedge \alpha \mid \triangle\epsilon,$$
$$\mathcal{L} \quad \stackrel{def}{=} \quad \varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [E]\varphi,$$

where p ranges over $\mathbb{P}$, ϵ ranges over $\mathcal{L}_{\mathsf{EQ}}$ and $E \in Int$. Operators $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined as usual abbreviations. The dual of $\Box$ is defined, as usual, as $\Diamond\varphi \stackrel{def}{=} \neg\Box\neg\varphi$ and the dual of $[E]$ is defined as $\langle E\rangle\varphi \stackrel{def}{=} \neg[E]\neg\varphi$.

Formula $\triangle\epsilon$ has to be read "the equational information ϵ is in the actual causal base". The modal formula $\Box\varphi$ has to be read "it is causally necessary that φ", while the modal formula $[E]\varphi$ has to be read "it will be causally necessary that φ, after the occurrence of the event E" or "if the event E occurred, φ would be necessarily true". In other words, $\Box$ is a modality of causal necessity, while $[E]$ is a modality of causal necessity *post intervention*.

The following notion of model is needed to provide a semantic interpretation of the formulas of the language $\mathcal{L}$.

Definition 5 (Equational model). *An equational model, or simply model, is a pair (S, U) such that $S \in U \subseteq \mathbf{S}_{Eq}$. The set of equational models is noted $\mathbf{M}_{eq}$.*

The component U is called *context* (or *universe*) of interpretation. $\mathcal{L}$-formulas are interpreted relative to a model, as follows. (We omit semantic interpretations for the Boolean connectives $\neg$, $\wedge$ and for $\top$ since they are defined in the usual way.)

Definition 6 (Semantic interpretation). *Let $(S, U) \in \mathbf{M}_{eq}$ with $S = (C, V)$. Then,*

$$
\begin{array}{lll}
(S,U) \models p & \textit{iff} & p \in V,\\
(S,U) \models \triangle\epsilon & \textit{iff} & \epsilon \in C,\\
(S,U) \models \Box\varphi & \textit{iff} & \textit{for all } S' \in U,\ \textit{if } S \equiv S'\\
 & & \textit{then } (S',U) \models \varphi,\\
(S,U) \models [E]\varphi & \textit{iff} & \textit{for all } S' \in U,\ \textit{if } S \Rightarrow^{E} S'\\
 & & \textit{then } (S',U) \models \varphi.
\end{array}
$$

Note that the modality $\Box\varphi$ could also be defined as an abbreviation of $[\emptyset]\varphi$ since they are logically equivalent. We decided to introduce it as a primitive for the sake of conceptual clarity, to clearly separate in the language unconditional causal necessity from conditional one.

Let us continue Example 2 of People v. Dlugash to illustrate the semantic interpretation of the modalities $\Box$ and $[E]$.

Example 3. *It is routine to verify that at the state $S_0 = (C_0, V_0)$, as defined in Example 1, it is necessarily the case that if Bush decides to shoot Geller then Geller will die. That is,*

$$(S_0, \mathbf{S}_{Eq}) \models \Box(Bd \rightarrow Di).$$

Moreover, at the state S_0, after having negatively intervened on the proposition Bs (e.g., by unloading Bush's gun), it is not necessarily the case that if Bush decides to shoot Geller then Geller will die. Indeed, the negative intervention on Bs breaks off the causal connection between Bush's action of deciding to shoot Geller and Bush's actions of shooting Geller and of killing him. That is,

$$(S_0, \mathbf{S}_{Eq}) \models \neg[\{Bs \leftrightarrow \bot\}](Bd \rightarrow Di).$$

3.3 Model Checking

In our framework, we can verify whether a causal model satisfies a causal property. The model checking procedure is schematically shown on Figure 1. It takes a finite equational state S_0, formed by a finite causal base C_0 and a valuation V_0, and a causal property $\varphi_0 \in \mathcal{L}$ as input. It then verifies whether $(S_0, \mathbf{S}_{Eq})$ satisfies φ_0 by translating the query $(S_0, \mathbf{S}_{Eq}) \models \varphi_0$ into a TQBF formula and then using a QBF solver on the translation.

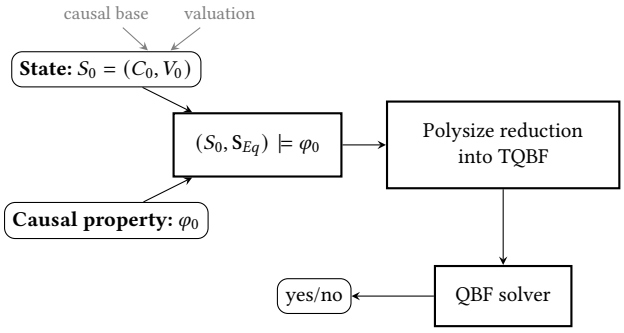

Figure 1: Causal model checker schema.

The computation time needed for the translation is a polynomial function on the size of the state and property combined. The technical details of this translation are shown on Appendix A. The details of the implementation are given in Section 5.7.

4 Causal Concepts

In this section, we introduce three causal notions: abductive explanation (AXp) [9, 13] also called sufficient reason [3] and PI-explanation [16], NESS (Necessary Element of a Sufficient Set) cause [17] and actual cause [7]. AXp is formulated using the causal necessity modality $\Box$ since it does not require the current causal model to be changed. Actual cause is formulated by means of the interventionist modality $[E]$, since it provides for counterfactual reasoning by changing the current causal model through an intervention. The formalization of AXp and actual cause given in this section relies on [4]. We consider two versions of NESS cause: a non-interventionist version (only requiring the "pure" causal necessity modality $\Box$) and an interventionist one. The former captures the essential aspects of Wright's original definition [17], while the latter is more in line with recent formalizations of NESS cause [1, 6]. Finally, we provide a stepwise approach to NESS and AXp, which may be connected to [1].

Some preliminary notation is needed before delving into the formalization of the three causal notions. Each of them involves two arguments: the *explanans* (or the *causing* fact) and the *explanandum* (or the *caused* fact). We assume the explanans is always a term, namely, a conjunction of literals in which a propositional variable can occur at most once, while the explanandum can be any formula φ of the language $\mathcal{L}$. We suppose the explanans λ and the explanandum φ do not share any atomic proposition. The set of terms is noted *Term* with elements noted $\lambda, \lambda', \ldots$. $Term_Z$ with $Z \subseteq \mathbb{P}$ denotes the set of terms built from the variables in Z. Given $\lambda, \lambda' \in Term$, with a slight abuse of notation, we write $\lambda' \subseteq \lambda$ (resp. $\lambda' \subset \lambda$) to mean that the set of literals appearing in λ' is a subset (resp. strict subset) of the set of literals appearing in λ.

4.1 Abductive Explanation

In line with [4] whose formalization of abductive explanation (AXp) is based on [9], we define an AXp to be a prime implicant of the explanandum that is actually true. Specifically, λ is an abductive explanation of φ if i) λ is actually true, ii) λ is an implicant of φ, and iii) there is no $\lambda' \subset \lambda$ such that λ' is also an implicant of φ. Conditions ii) and iii) define the concept of prime implicant.

Definition 7 (Abductive explanation). *Let $S = (C, V) \in \mathbf{S}_{Eq}$, $U \subseteq \mathbf{S}_{Eq}$ and $\lambda \in Term_{\mathbb{P}(C)}$. We say λ is an abductive explanation (AXp) of φ at (S, U) if the following holds:*

$$(S, U) \models \lambda \wedge \mathsf{PImp}(\lambda, \varphi)$$

with

$$\mathsf{PImp}(\lambda, \varphi) \stackrel{def}{=} \Box(\lambda \rightarrow \varphi) \wedge \bigwedge_{\lambda' \subset \lambda} \neg\Box(\lambda' \rightarrow \varphi).$$

Note that the definition of AXp only requires the "pure" causal necessity modality $\Box$.

4.2 NESS Cause

Inspired by [17], we say that a term λ is a NESS cause of φ if it is part of a larger term which i) is actually true, ii) is causally sufficient to make φ true, and iii) is no longer causally sufficient to make φ true in the absence of λ.

Definition 8 (NESS cause). *Let $S = (C, V) \in \mathbf{S}_{Eq}$, $U \subseteq \mathbf{S}_{Eq}$ and $\lambda \in Term_{\mathbb{P}(C)}$. We say λ is a NESS cause of φ at (S, U) if the following holds:*

$$(S, U) \models \lambda \wedge \bigvee_{\substack{\lambda' \in Term_{\mathbb{P}(C)}: \\ \lambda \cap \lambda' = \emptyset}} \big(\lambda' \wedge \mathsf{NecSuff}(\lambda, \lambda', \varphi)\big),$$

with

$$\mathsf{NecSuff}(\lambda, \lambda', \varphi) \stackrel{def}{=} \Box\big((\lambda \wedge \lambda') \rightarrow \varphi\big) \wedge \bigwedge_{\lambda'' \subset \lambda} \neg\Box\big((\lambda'' \wedge \lambda') \rightarrow \varphi\big).$$

We may also define an interventionist variant of NESS cause which is analogous to the notion of 'direct NESS cause' given in [1, Definition 5] and 'cause according to the causal NESS test' given in [6, Definition 5.3]. The crucial difference between plain NESS of Definition 8 and the following notion of interventionist NESS cause is that in the former causal sufficiency of $\lambda \wedge \lambda'$ relative to φ is expressed through the causal necessity modality (i.e., $\Box\big((\lambda \wedge \lambda') \rightarrow \varphi\big)$), while in the latter it is expressed through the interventionist modality (i.e., $[E_{\lambda \wedge \lambda'}]\varphi$).

Definition 9 (Interventionist NESS cause). *Let $S = (C, V) \in \mathbf{S}_{Eq}$, $U \subseteq \mathbf{S}_{Eq}$ and $\lambda \in Term_{\mathbb{P}(C)}$. We say λ is an intervenionist NESS cause of φ at (S, U) if the following holds:*

$$(S, U) \models \lambda \wedge \bigvee_{\substack{\lambda' \in Term_{\mathbb{P}(C)}: \\ \lambda \cap \lambda' = \emptyset}} \big(\lambda' \wedge \mathsf{IntervNecSuff}(\lambda, \lambda', \varphi)\big),$$

with

$$\mathsf{IntervNecSuff}(\lambda, \lambda', \varphi) \stackrel{def}{=} [E_{\lambda \wedge \lambda'}]\varphi \wedge \bigwedge_{\lambda'' \subset \lambda} \neg[E_{\lambda'' \wedge \lambda'}]\varphi$$

where for every term λ we define:

$$E_\lambda \stackrel{def}{=} \big\{p \leftrightarrow \top : p \subseteq \lambda\big\} \cup \big\{p \leftrightarrow \bot : \neg p \subseteq \lambda\big\}.$$

The only difference between our definition of interventionist NESS and Beckers' definition and Halpern's definition given respectively in [1, Definition 5] and [6, Definition 5.3] is that their definitions are relative to a specific actual assignment of the exogenous variables while our definition is not, since we universally quantify over all possible assignments of the (exogenous and endogenous) variables that are causally compatible with the actual state after the interventions. Note that our notion of interventionist NESS implies Beckers' and Halpern's notions since our notion of sufficiency is relative to all possible assignments of the endogenous variables, while theirs is relative to the actual assignment. Also note that Beckers' and Halpern's notions are definable in our language.

4.3 Actual Cause

In line with [4] whose formalization of actual cause is based on [7], we say that λ is an actual cause of φ if i) both λ and φ are true, ii) it is possible to intervene on the endogenous variables in λ, while fixing by intervention the values of some endogenous variables that are not in λ, such that if the value of the exogenous variables do not change, φ will be necessarily false, and iii) there is no $\lambda' \subset \lambda$ for which ii) holds.

Definition 10 (Actual cause). *Let $S = (C, V) \in \mathbf{S}_{Eq}$, $U \subseteq \mathbf{S}_{Eq}$ and $\lambda \in Term_{end(S)}$. We say λ is an actual cause of φ at (S, U) if the following holds:*

$$(S, U) \models \lambda \wedge \varphi \wedge \mathsf{But}(S, \lambda, \varphi) \wedge \bigwedge_{\lambda' \subset \lambda} \neg\mathsf{But}(S, \lambda', \varphi),$$

where

$$\mathsf{But}(S, \lambda, \varphi) \stackrel{def}{=} \bigvee_{\substack{Z \subseteq end(S), \\ Z \cap \mathbb{P}(\lambda) = \emptyset, \\ E \in Int_{\mathbb{P}(\lambda)}, \\ E' \in Int_Z}} \Big(\widehat{E'} \wedge [E \cup E']\big(\lambda_S^{exo} \rightarrow \neg\varphi\big)\Big)$$

and

$$\lambda_S^{exo} = \bigwedge_{p \in exo(S) \cap V} p \wedge \bigwedge_{p \in exo(S) \setminus V} \neg p.$$

Note that, like interventionist NESS of Definition 9, the definition of actual cause requires the interventionist modality $[E]$.

4.4 Stepwise AXp and NESS

As we will show at a later stage, AXp (Definition 7) as well as non-interventionist and interventionist NESS cause (Definition 8 and 9) are unable to capture preemption, since both the preemptor and preempted fact are considered to be causes. The limitation of AXp and standard NESS in capturing preemption is due to the fact they do not identify what elements would block the cause from producing the effect when other causes are present which produce anyway the effect. One way to overcome this limitation, in line with Beckers' idea of counterfactual NESS cause [1, Definition 8] and [18], consists in defining notions of stepwise AXp or NESS cause. For a literal l to be a cause of another literal l', it must exist a sequence (or causal path) of literals $l_1, \ldots, l_n$ such that $l = l_1$, $l' = l_n$,

and for every $1 \leq k < n$ i) the variable of the literal l_k is connected to the variable of the literal l_{k+1} in the causal graph induced by the actual causal base,[3] ii) l_k is an AXp or NESS cause of l_{k+1} in the sense of Definitions 7 and 8 respectively. Let us first define stepwise AXp that, as far as we know, has never been considered in the literature. It is the first refinement of the notion of abductive explanations that is able to capture preemption.

Definition 11 (Stepwise AXp). *We say literal l is a stepwise AXp of literal l' at (S, U) iff there exists $l_1, \ldots, l_n$ with $l = l_1$, $l' = l_n$ such that $l = l_1$, $l' = l_n$, and for every $1 \leq k < n$*

$$(S, U) \models \bigvee_{\mathbb{P}(l_{k+1}) \leftrightarrow \omega \in C : \mathbb{P}(l_k) \in \mathbb{P}(\omega)} \triangle\big(\mathbb{P}(l_{k+1}) \leftrightarrow \omega\big)$$

and l_k is an AXp of l_{k+1} at (S, U).

The following definition introduces stepwise NESS cause. For the sake of simplicity we only consider the non-interventionist variant since it is sufficient to capture preemption. Defining the interventionist variant would be straightforward by simply replacing 'NESS cause' by 'interventionist NESS cause' in the definition below.

Definition 12 (Stepwise NESS cause). *We say literal l is a stepwise NESS cause of literal l' at (S, U) iff there exists $l_1, \ldots, l_n$ with $l = l_1$, $l' = l_n$ such that $l = l_1$, $l' = l_n$, and for every $1 \leq k < n$*

$$(S, U) \models \bigvee_{\mathbb{P}(l_{k+1}) \leftrightarrow \omega \in C : \mathbb{P}(l_k) \in \mathbb{P}(\omega)} \triangle\big(\mathbb{P}(l_{k+1}) \leftrightarrow \omega\big)$$

and l_k is a NESS cause of l_{k+1} at (S, U).

The previous definitions of stepwise AXp and NESS have however one limitation. They only work for literals and it is not clear how to generalize them to arbitrary formulas either for the cause or for the caused fact.

5 Causation in Legal Cases: Scenario Analysis

In this section, we present cause-in-fact analysis in selected legal cases, as a testbed for the concepts of causation introduced above. We use an implemented causal model checker, parametrised to such concepts to ask causal queries. The queries are run to check for the six causal concepts: abductive explanation (AXp), NESS (N), interventionist NESS (N_I), actual cause (AC), stepwise abductive explanation (AXp_S), and stepwise NESS (N_S). The answers to all relevant queries are shown in Table 1 (page 8).

Each legal case presented in the sequel has been modelled using the formalism presented above. In the formalisation of each case below, we list the relevant propositional variables of the problem and their meanings, as well as the description of the actual state given as input to the model checker. In each case, the actual state is $S_0 = (C_0, V_0)$ with C_0 being the actual causal base and V_0 the set of variables that are actually true.

[3]The procedure for extracting a causal graph from a causal base C is as follows: i) draw one node for every variable appearing in $\mathbb{P}(C)$, ii) for every equational formula $p \leftrightarrow \omega$ in C and for every variable q occurring in ω draw an edge from q to p.

5.1 Asbestos and Lung Cancer (Overdetermination)

Fairchild v. Glenhaven (2002)[4] is a seminal case in the UK concerning asbestos exposure from several sources (linked to various companies) and lung cancer. Mr. Fairchild worked for three construction companies for several years in the 1960s, handling asbestos materials. In 1996 he died of lung cancer. Expert witnesses attested that the exposure to asbestos from two of the companies was sufficient to cause his cancer. Subsequently, all three companies were held liable for his death. We formalise it as follows:

- Co_i: Farchild worked for a company i (exogenous).
- Ex_i: Farchild has been exposed to asbestos in i.
- Ca: Farchild contracted lung cancer.
- Di: Farchild died.

$C_0 = \{Ca \leftrightarrow ((Ex_1 \wedge Ex_2) \vee (Ex_1 \wedge Ex_3) \vee (Ex_2 \wedge Ex_3)),$
$Di \leftrightarrow Ca, Ex_1 \leftrightarrow Co_1, Ex_2 \leftrightarrow Co_2, Ex_3 \leftrightarrow Co_3\}$

$V_0 = \{Co_1, Co_2, Co_3, Ex_1, Ex_2, Ex_3, Ca, Di\}$.

Note that the difference in the answers (see Table 1) concerns the treatment of one of the literals (Ex_1, Ex_2 and Ex_3) as a cause, where only the NESS variants identify them correctly.

5.2 The Tobacco Case (Confounding)

In the context of the infamous US tobacco cases (where companies were accused of lying to consumers about the negative effects of tobacco), the companies claimed that confounders could explain the correlation between smoking and lung cancer, and thus undermine the claim that there was a causal relation between the two (for a discussion see [15, Ch. 5]).

For instance, a genetic predisposition favouring both tobacco consumption and lung cancer could indicate a correlation without causation. This causal model was rejected by the judges, but if it was accepted, causation would be excluded. We consider whether our concepts of causality provide the correct answer for this model (no causality). We formalise it as follows.

- Mk: Person has genetic makeup (exogenous).
- Pr: Person has genetic predisposition.
- Sm: Person smokes tobacco.
- Lu: Person has lung cancer.

$C_0 = \{Lu \leftrightarrow Pr, Sm \leftrightarrow Pr, Pr \leftrightarrow Mk\}$

$V_0 = \{Mk, Pr, Sm, Lu\}$

Interventionist NESS, actual cause, stepwise NESS and stepwise AXp all provide the correct answer, namely that smoking (Sm) does not cause lung cancer (Lu). The other concepts fail since they are confounded by smoking. No concept considers the conjunction $Sm \wedge Pr$ a cause, which is correct, since it is not minimal (Pr alone is sufficient).

5.3 Double Shooting, Attempted Crime (Preemption)

People v. Dlugash (1997) is a criminal case where the defendant attempted to murder someone who was already dead. We introduced the facts of the case in Section 2 and used it as our running example. Thus here we only repeat the atoms and provide the corresponding formalisation.

[4]*Fairchild v Glenhaven Funeral Services Ltd* [2002] UKHL 22.

Bd/Dd: Bush/Dlugash decides to shoot Geller (exogenous).
Bs/Ds: Bush/Dlugash shoots Geller.
Bk/Dk: Bush/Dlugash kills Geller.
Di: Geller dies.

$C_0 = \{Di \leftrightarrow Bk \vee Dk, Bk \leftrightarrow Bs, Dk \leftrightarrow (Ds \wedge \neg Bk), Bs \leftrightarrow Bd, Ds \leftrightarrow Dd\}$

$V_0 = \{Di, Bk, Bs, Ds, Bd, Dd\}$

In this example, only actual causation, stepwise NESS and stepwise AXp provide the right answers (i.e., only Bush's shot is a cause of Geller's death). The other approaches are unable to deal with preemption. All approaches correctly answer 'no' to the conjunctive query, since the conjunction is not minimal (Bush's action alone is enough).

5.4 Childhood Leukaemia (Omission)

This Italian case (Cassazione penale, 2023)[5] concerns parents' responsibility for the death of their child, who was diagnosed with acute lymphoblastic leukaemia. Doctors strongly recommended chemotherapy, but the parents refused the treatment and the child died soon after. The parents were held responsible under civil law for causing the death of the child. We formalise it as follows.

Le: Child has leukaemia (exogenous).
Di: Child dies.
Ch: Child receives chemotherapy.
Co: Parents consent.
De: Parents decide to consent (exogenous).

$C_0 = \{Di \leftrightarrow (\neg Ch \wedge Le), Ch \leftrightarrow Co, Co \leftrightarrow De\}$

$V_0 = \{Di, Le\}$

In this case, the parents failure to consent is the cause of the death of the child, according to all causal concepts except abductive explanation. In other words, omission is identified as a cause.

5.5 Car Accident (Preemptive Negative Causation)

The *Saunders System Birmingham v. Adams* (1928)[6] concerns a road accident where a car runs into a motorist. The driver did not push the brake pedal, but it was later shown that the brakes were defective so even if the driver had tried to brake the accident would have happened anyway. Only the driver was considered to have caused the crash. Following [18] we formalise it as follows.

Ac: Accident happens.
Fa: Brakes fail.
Pu: Driver pushes brake pedal.
Ma: Brakes malfunction.
Df: Brakes are defective (exogenous).
De: Driver decides to brake (exogenous).

$C_0 = \{Ac \leftrightarrow Fa \vee \neg Pu, Fa \leftrightarrow Ma \wedge Pu, Pu \leftrightarrow De, Ma \leftrightarrow Df\}$

$V_0 = \{Ac, Df, Ma\}$

According to [18], the driver's failure to push the brake pedal preempted the brake defect to cause the brake failure, so that only the driver's inaction is the cause. As expected, the right result is provided by actual cause, stepwise NESS and stepwise AXp which can deal with preemption.

[5] Case 12124/2023, Cassazione Penale.
[6] *Saunders System Birmingham Co. v. Adams*, 61 A.L.R. 1333 (Ala. 1928).

5.6 War Crime (Necessary Causation or Ennoblement)

In the following, we consider a war crime discussed in the literature [1] in which both a proposition and its negation can produce the effect. The sergeant requests the soldier to shoot a prisoner, and tells the soldier: "if you do not shoot the prisoner, I will do it". The solder complies. We formalise it as follows.

Di: Prisoner dies.
SoS/SeS: Soldier/Sergeant shoots prisoner.
SoH/SeH: Soldier/Sergeant hits prisoner.
SoD: Soldier decides to shoot (exogenous).

$C_0 = \{Di \leftrightarrow SoH \vee SeH, SeH \leftrightarrow SeS, SoH \leftrightarrow SoS, SeS \leftrightarrow \neg SoS, SoS \leftrightarrow SoD\}$

$V_0 = \{Di, SoH, SoS, SoD\}$

In this case, the conclusion that SoS is a cause of Di (which seems intuitively right) is given by both actual cause and interventionist NESS. SoS is not a cause according to the other concepts, since Di is necessary (it happens both if SoS and if $\neg SoS$). Note that, if the soldier does not shoot the prisoner, (i.e., $V_0 = \{Di, SeS\}$), this omission ($\neg SoS$) would be a cause of the prisoner's death by the same concepts.

This would be an instance of ennoblement in a strict sense [19]: while the shooting would have directly caused the death, by not shooting the soldier indirectly determines the same outcome, by inducing (ennobling) causation by the sergeant's action.

5.7 Implementation

In order to verify the feasibility of the theoretical framework, we implemented a symbolic model checker, which uses the translation into TQBF. The resulting TQBF is then translated into a binary decision diagram (BDD). The program is implemented in Haskell and the BDD library used is HasCacBDD. It was compiled with GHC 9.4.8 and executed in a MacBook Air with a 1.6 GHz Dual-Core Intel Core i5 processor and 16 GB of RAM, running macOS Sonoma 14.7.1.

The computation times in Table 2 were obtained by calculating the average of the elapsed time of 55 runs of the compiled program on each query. More runs could be performed, but the variation on computation time between each run is small, since the program is completely deterministic. The maximum standard deviation observed is 0.5328 seconds (less than 6% of the expected value of 9.6061) on the interventionist NESS column of the last Asbestos line.

Only queries using NESS, interventionist NESS and stepwise NESS needed more than one second to complete. Interventionist NESS was the slowest query in every line of the table. The slowest of all being the interventionist NESS on the last line of Asbestos, which took nearly 10 seconds. Queries to concepts based on NESS need more time to complete because they use the modal operators twice. In other words, the formulas in NESS concepts are more complicated than on the other concepts. It is also not surprising that the concepts whose queries needed less time to complete were either AXp or AXp_S, the less complicated ones. These figures show that the concrete legal cases we tested can be model checked in reasonable time.

Table 1: Answers to the causal queries. For example, the first 'no' in the table means that Ex_1 is not a cause of Di in the Asbestos cause according to AXp. The "intuitively correct" answers are in bold face.

	AXp	N	N_I	AC	AXp_S	N_S
Asbestos (overdet.): causes for Fairchild's death (Di)						
Ex_1	no	**yes**	**yes**	no	no	**yes**
Ex_2	no	**yes**	**yes**	no	no	**yes**
Ex_3	no	**yes**	**yes**	no	no	**yes**
$Ex_1 \wedge Ex_2$	**yes**	**yes**	**yes**	**yes**	n/a	n/a
$Ex_1 \wedge Ex_3$	**yes**	**yes**	**yes**	**yes**	n/a	n/a
$Ex_2 \wedge Ex_3$	**yes**	**yes**	**yes**	**yes**	n/a	n/a
$Ex_1 \wedge Ex_2 \wedge Ex_3$	**no**	**no**	**no**	**no**	n/a	n/a
Tobacco (confounding): causes for lung cancer (Lu)						
Sm	yes	yes	**no**	**no**	**no**	**no**
Pr	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
$Sm \wedge Pr$	**no**	**no**	**no**	**no**	n/a	n/a
Double shooting (preempt.): causes for Geller's death (Di)						
Ds	yes	yes	yes	**no**	**no**	**no**
Bs	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
$Bs \wedge Ds$	**no**	**no**	**no**	**no**	n/a	n/a
Leukaemia (omission): causes for the child's death (Di)						
$\neg Co$	no	**yes**	**yes**	**yes**	no	**yes**
Car accident (neg. preempt.): causes for the car accident (Ac)						
$\neg Pu$	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
Ma	yes	yes	yes	**no**	**no**	**no**
$\neg Pu \wedge Ma$	**no**	**no**	**no**	**no**	n/a	n/a
War crime (ennobl.): causes for the prisoner's death (Di)						
SoS	no	no	**yes**	**yes**	no	no
$\neg SeS$	**no**	**no**	**no**	**no**	**no**	**no**
War crime (cont.): when the soldier does not shoot ($\neg Sos$)						
$\neg SoS$	no	no	**yes**	**yes**	no	no
SeS	no	no	**yes**	**yes**	no	no

Table 2: Computation times in seconds for the complete set of queries.

	AXp	N	N_I	AC	AXp_S	N_S
Asbestos (overdet.): causes for Fairchild's death (Di)						
Ex_1	0.0686	2.0248	6.1635	0.1732	0.0022	3.9160
Ex_2	0.0003	1.9231	6.1249	0.1758	0.0006	3.8246
Ex_3	0.0003	1.9302	6.1191	0.1754	0.0005	3.9184
$Ex_1 \wedge Ex_2$	0.0005	2.3551	7.4601	0.2715	n/a	n/a
$Ex_1 \wedge Ex_3$	0.0005	2.3792	7.3681	0.2638	n/a	n/a
$Ex_2 \wedge Ex_3$	0.0005	2.3376	7.3725	0.2590	n/a	n/a
$Ex_1 \wedge Ex_2 \wedge Ex_3$	0.0025	3.0866	**9.6061**	0.4100	n/a	n/a
Tobacco (confounding): causes for lung cancer (Lu)						
Sm	0.0002	0.0107	0.0242	0.0080	0.0000	0.0000
Pr	0.0001	0.0105	0.0233	0.0076	0.0001	0.0106
$Sm \wedge Pr$	0.0002	0.0120	0.0289	0.0113	n/a	n/a
Double shooting (preemption): causes for Geller's death (Di)						
Ds	0.0003	0.5991	1.6786	0.1705	0.0005	1.1898
Bs	0.0018	0.6065	1.6643	0.1726	0.0021	1.1961
$Ds \wedge Bs$	0.0005	0.7539	2.0344	0.2613	n/a	n/a
Leukaemia (omission): causes for child's death (Di)						
$\neg Co$	0.0002	0.0369	0.1025	0.0082	0.0004	0.1128
Car accident (negative preemption): causes for the car accident (Ac)						
$\neg Pu$	0.0002	0.1605	0.4319	0.0413	0.0002	0.1625
Ma	0.0002	0.1614	0.4299	0.0392	0.0004	0.3228
$\neg Pu \wedge Ma$	0.0020	0.1962	0.5015	0.0595	n/a	n/a
War crime (ennoblement): causes for the prisoner's death (Di)						
SoS	0.0003	0.1922	0.4561	0.1673	0.0039	1.9323
$\neg SeS$	0.0003	0.1918	0.4446	0.1621	0.0023	0.5812
…when the soldier does not shoot ($\neg SoS$)						
$\neg SoS$	0.0003	0.1993	0.4597	0.1674	0.0039	2.0071
SeS	0.0003	0.2005	0.4613	0.1701	0.0008	0.6080

6 Discussion

The six concepts of causality presented above behave differently (see Table 1), and none of them gives the "intuitively right" answer for legal responsibility across all cases. We summarise the strengths and weaknesses of each theory below.

In the overdetermination scenario, the correct identification of causes includes any pair combinations and any single literal (e.g., each individual company involved in the asbestos exposure). While abductive explanation and actual causation correctly identify the pairs, only the NESS variants also recognise single contributors as causes, aligning with the court's reasoning. In confounding cases, interventionist NESS and actual causation are the most accurate, as they identify the single cause and reject non-minimal pairs. Confounding challenges approaches based on logical necessity, such as plain NESS and AXp. Preemption scenarios present difficulties for AXp, plain NESS, and interventionist NESS, as these methods fail to exclude the preempted action as a cause. Actual causation and stepwise approaches, however, handle preemption more effectively. Instances of omission are recognised as causes by all NESS variants and actual causation, while abductive explanation fails to capture omission. Specifically, AXp struggles in these scenarios because it relies on sufficiency, requiring both the event and the omission in conjunction to qualify as a cause. Finally, the precarious case of ennoblement is best addressed by interventionist approaches — actual causation and interventionist NESS — while other methods find ennoblement events causally ambiguous. Overall, stepwise NESS performs well in most scenarios involving single literals but requires interventionist supplementation for ennoblement cases, which will be explored in future work.

This diversity highlights the value of these concepts as a flexible analytical toolbox for legal reasoners. Thus, we may conclude that a suitable formal concept for cause-in-fact, in legal (but we also believe in other) contexts can be provided by selecting the appropriate method in the toolbox we have described, being aware of the limitation of each method. By selecting the method best suited to the specific scenario and understanding the limitations of each approach, legal reasoning can achieve greater transparency, coherence, and precision. Moreover, such analysis requires reasoners to uncover hidden assumptions and reasoning patterns, promoting transparency and encouraging critical evaluation of legal decisions.

7 Conclusion

In this paper, we compared how formal causal approaches handle causal complexities found in legal cases. We recognise that the notion of cause-in-fact does not encompass all aspects relevant to determining legal liability for harm. These include considerations such as the normality or foreseeability of causal effects, the mental states of actors (e.g., intention, recklessness, or negligence), the presence of justifications and exemptions (e.g., necessity, self-defense, or incapacity), and the evaluation of evidence, including presumptions and burdens of proof. However, as argued by Wright [18], causation-in-fact remains a crucial element in establishing liability. A rigorous approach to cause-in-fact is critical for addressing complex issues like overdetermination, negative causation, preemption, confounding, and ennoblement. We believe that the logical framework and the model checking algorithm presented in the paper can help i) legal practitioners to compute the causal attributions provided by a notion of cause-in-fact in a specific legal case, and ii) legal theorists to better understand the pros and cons of each notion of cause by automatically verifying its causal attributions in a large body of legal cases. Furthermore, examining the interplay between cause-in-fact and legally significant causal considerations represents a promising avenue for future research, particularly in a period where AI liability is a growing focus of practical interest.

This paper contributes to the growing body of research on the formal analysis of causation, with a particular focus on its applications in practical areas such as law. This is a topic little investigated within AI & law research (see, however, [10–12], for partially formalised accounts). The novel contributions include the formalisation of approaches, the introduction of stepwise NESS and stepwise AXp variants, and their technical implementation for analysing real legal cases. To our knowledge, it is also, the first attempt to systematically compare these approaches using real legal cases. In future work, we aim to explore the causal complexities in AI liability cases and the explanatory potential of analyses using the causal model checker.

Acknowledgments

This work was partially supported by the following projects: CompuLaw (Computable Law), funded by the ERC under the Horizon 2020 (Grant Agreement N. 833647). "FAIR - Future Artificial Intelligence Research": Spoke 8 and Spoke 1 (CAI4DSA action) under the European Commission's NextGeneration EU programme, PNRR – M4C2 – Inv. 1.3, Partenariato Esteso (PE00000013). ANR AI Chair project "Responsible AI" (grant number ANR-19-CHIA-0008).

References

[1] S. Beckers. 2021. The Counterfactual NESS Definition of Causation. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-2021)*. AAAI Press, 6210–6217.
[2] O. Beyersdorff, M. Janota, F. Lonsing, and M. Seidl. 2021. Quantified Boolean Formulas. In *Handbook of Satisfiability*, A. Biere, M. Heule, H. van Maaren, and T. Walsh (Eds.). Vol. 336. 1177–1221.
[3] A. Darwiche and A. Hirth. 2020. On the Reasons Behind Decisions. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020) (Frontiers in Artificial Intelligence and Applications, Vol. 325)*. IOS Press, 712–720.
[4] T. de Lima and E. Lorini. 2024. Model Checking Causality. In *Proceedings of the Thirty-Thirth International Joint Conference on Artificial Intelligence, (IJCAI 2024)*. ijcai.org, 3324–3332.
[5] N. Hall and L. A. Paul. 2003. Causation and Preemption. In *Philosophy of science today*, P. Clark and K. Hawley (Eds.). Oxford University Press, 100–130.
[6] J. Y. Halpern. 2008. Defaults and Normality in Causal Structures. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR 2008)*. AAAI Press, 198–208.
[7] J. Y. Halpern. 2015. A Modification of the Halpern-Pearl Definition of Causality. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*. AAAI Press, 3022–3033. http://ijcai.org/Abstract/15/427
[8] H. L. A. Hart and T. Honoré. 1959. *Causation in Law*. Clarendon.
[9] A. Ignatiev, N. Narodytska, and J. Marques-Silva. 2019. Abduction-Based Explanations for Machine Learning Models. In *Proceedings of the The Thirty-Third AAAI Conference on Artificial Intelligence (AAA-19)*. 1511–1519.
[10] J. Lehmann, J. Breuker, and B. Brouwer. 2004. Causation in AI and Law. *Artif. Intell. Law* 12 (12 2004), 279–315.
[11] R. Liepina, G. Sartor, and A. Wyner. 2015. Causal models of legal cases. In *International Workshop on AI Approaches to the Complexity of Legal Systems*. Springer, 172–186.
[12] R. Liepina, G. Sartor, and A. Wyner. 2020. Arguing about causes in law: a semi-formal framework for causal arguments. *Artificial Intelligence and Law* 28 (2020), 69–89.
[13] X. Liu and E. Lorini. 2023. A unified logical framework for explanations in classifier systems. *Journal of Logic and Computation* 33, 2 (2023), 485–515.
[14] E. Lorini. 2023. A Rule-Based Modal View of Causal Reasoning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI 2024)*. ijcai.org, 3286–3295.
[15] J. Pearl and D. Mackenzie. 2018. *The Book of Why*. Basic Books.
[16] A. Shih, A. Choi, and A. Darwiche. 2018. A Symbolic Approach to Explaining Bayesian Network Classifiers. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*. 5103–5111.
[17] R. W. Wright. 1988. Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts. *Iowa Law Review* 73 (1988), 1001–1077.
[18] R. W. Wright. 2011. The NESS account of natural causation: A response to criticisms. *Perspectives on causation* 285 (2011), 305.
[19] S. Yablo. 2010. Advertisement for a Sketch of an Outline of a Proto Theory of Causation. In *Things: Papers on Objects, Events, and Properties*. Oxford University Press, 98–116.

A Appendix

In this section, we explain the technical details governing the model checking procedure.

First, note that, according to Definition 5, the U-component of a model can be any (possibly infinite) set of states with no restriction on the equational information that is included in a causal base. From a practical point of view, it is useful to restrict to models in which causal bases are constructed from a finite repository of causal information $\Gamma \subset \mathcal{L}_{\text{EQ}}$.

Definition 13 (Γ-model). *Let a finite $\Gamma \subset \mathcal{L}_{\text{EQ}}$ be given. The model (S, U) is said to be a Γ-model if $S \in U = \mathbf{S}_\Gamma$, with $\mathbf{S}_\Gamma = \{(C, V) \in \mathbf{S}_{Eq} : C \subseteq \Gamma\}$.*

Second, notice that interventions can only change an equational state into another equational state. That is, if S_0 is an equational state and $S \Rightarrow^E S'$ then S' is an equational state too. Therefore, by induction on the structure of the formula φ, we can easily prove the following equivalence, under the assumption that S_0 is an equational state:

$$(S_0, \mathbf{S}_{Eq}) \models \varphi \text{ iff } (S_0, S) \models \varphi \tag{1}$$

Furthermore, it has been shown in [4, Theorem 7] that a vocabulary Γ_{S_0,φ_0} can be constructed from S_0 and φ_0 in such a way that:

$$(S_0, S) \models \varphi \text{ iff } (S_0, \mathbf{S}_{\Gamma_{S_0,\varphi_0}}) \models \varphi \tag{2}$$

Therefore, by (1) and (2), we have that, if S_0 is an equational state, then:

$$(S_0, \mathbf{S}_{Eq}) \models \varphi \text{ iff } (S_0, S_{\Gamma_{S_0,\varphi_0}}) \models \varphi$$

In our implementation, we use the latter result to automatically generate such a vocabulary and define model checking in our setting as follows:

Definition 14 (Model checking).

input: *a finite state* $S_0 \in \mathrm{S}_{Eq}$*, and a formula* $\varphi_0 \in \mathcal{L}$.
output: *yes if* $(S_0, \mathrm{S}_{Eq}) \models \varphi_0$*; no otherwise.*

To be able to provide an efficient algorithm, we use a translation to Quantified Boolean Formulas (QBF). The first step of the procedure generates the vocabulary $\Gamma = \Gamma_{S_0,\varphi_0}$ from S_0 and φ_0. After that, let Γ and φ_0 be given. The set W of relevant sub-formulas is defined as $W = \mathbb{P}(\varphi_0) \cup \mathbb{P}(\Gamma) \cup \{\triangle\epsilon \mid \epsilon \in \Gamma\}$. A set of fresh propositional variables is defined for each state. That is, for each $S \in \mathrm{S}_\Gamma$, we have $W_S = \{x_{\varphi,S} \mid \varphi \in W\}$, where $x_{\varphi,S}$ is a fresh propositional variable not appearing in Γ or φ_0.

The translation of the model checking problem $(S, \mathrm{S}) \models \varphi$ to QBF is recursively defined as follows.

Definition 15 (Translation).

$$\begin{aligned}
\mathrm{tr}(\top, S) &= \top \\
\mathrm{tr}(p, S) &= x_{p,S} \\
\mathrm{tr}(\triangle\epsilon, S) &= \begin{cases} x_{\triangle\epsilon,S}, & \text{if } \epsilon \in \Gamma \\ \bot, & \text{otherwise} \end{cases} \\
\mathrm{tr}(\neg\varphi, S) &= \neg\, \mathrm{tr}(\varphi, S) \\
\mathrm{tr}(\varphi_1 \wedge \varphi_2, S) &= \mathrm{tr}(\varphi_1, S) \wedge \mathrm{tr}(\varphi_2, S) \\
\mathrm{tr}(\Box\varphi, S) &= \forall W_{S'} (T^{\emptyset}_{(S,S')} \rightarrow \mathrm{tr}(\varphi, S')) \\
\mathrm{tr}([E]\varphi, S) &= \forall W_{S'} (T^{E}_{(S,S')} \rightarrow \mathrm{tr}(\varphi, S'))
\end{aligned}$$

with

$$T^{E}_{(S,S')} \stackrel{def}{=} \bigwedge_{\epsilon\in\Gamma} g(\epsilon, E, S, S'),$$

and

$$g(\epsilon, E, S, S') = \begin{cases} \neg x_{\triangle\epsilon,S'}, \\ \quad \text{if } \epsilon \in \left(C \cap \bigcup_{p\in\mathbb{P}(E)} \mathcal{L}_{\mathrm{EQ}}(p)\right) \setminus E \\ x_{\triangle\epsilon,S'} \wedge \mathrm{tr}(\epsilon, S'), \ \text{if } \epsilon \in E \\ (x_{\triangle\epsilon,S} \leftrightarrow x_{\triangle\epsilon,S'}) \wedge \\ (x_{\triangle\epsilon,S} \rightarrow \mathrm{tr}(\epsilon, S')), \ \text{otherwise} \end{cases}$$

Intuitively, each fresh variable $x_{p,S}$ encodes the membership of p to V, whereas $x_{\triangle\epsilon,S}$ encodes the membership of ϵ to C. This means that each state $S \in \mathrm{S}_\Gamma$ can be associated with an element of 2^{W_S}, namely $\{x_{p,S} \mid p \in V\} \cup \{x_{\triangle\epsilon,S} \mid \epsilon \in C\}$. Therefore, we also have that formula $T^{E}_{(S,S')}$ encodes the membership of (S, S') to the relation $\Rightarrow^E$, as follows. If ϵ is removed from the base of S, then $\triangle\epsilon$ is set to false in S'. Similarly, if ϵ is added to the base of S, then $\triangle\epsilon$ is set to be true in S'. In the latter case, one must also make sure that S' is a state. Therefore, $\mathrm{tr}(\epsilon, S')$ is also set to be true at S'. In the third case, if ϵ is not added nor removed, then its truth value must remain the same in S'. In addition, in the case where it is true, ϵ must also be satisfied by S'. Also note that the formulas $\triangle\epsilon$ which are not in W are considered to be false in the model. Therefore, the translation replaces them with $\bot$. The theorem below shows how the translation is used to encode model checking.

Theorem 16. *Let* $S = (C, V)$ *be a state in* S_Γ. $(S, \mathrm{S}_\Gamma) \models \varphi$ *if and only if* $\models_{QBF} \forall W_S (D_S \rightarrow \mathrm{tr}(\varphi, S))$*, where:*

$$D_S = \bigwedge_{\epsilon\in C} x_{\triangle\epsilon,S} \wedge \bigwedge_{\epsilon\in W\setminus C} \neg x_{\triangle\epsilon,S} \wedge \bigwedge_{p\in V} x_{p,S} \wedge \bigwedge_{p\in W\setminus V} \neg x_{p,S}$$

Proof. Let $val = \bigcup_{S\in\mathrm{S}_\Gamma} val_S$, where, for each state $S = (C, V) \in \mathrm{S}_\Gamma$, $val_S = \{x_{p,S} \mid p \in V\} \cup \{x_{\triangle\epsilon,S} \mid \epsilon \in C\}$. In the sequel, the set val is used as a QBF model.

First, we show that $(S, \mathrm{S}_\Gamma) \models \varphi$ iff $val \models \mathrm{tr}(\varphi, S)$ by induction on the structure of φ.

There are three cases in the induction base:

- Let $\varphi = \top$. $(S, \mathrm{S}_\Gamma) \models \top$, iff $val \models \top$, iff $val \models \mathrm{tr}(\top, S)$.
- Let $\varphi \in \mathbb{P}$. $(S, \mathrm{S}_\Gamma) \models p$ iff $p \in V$, iff $x_{p,S} \in val_S$, iff $val_S \models x_{p,S}$, iff $val \models \mathrm{tr}(p, S)$.
- Let $\varphi = \triangle\epsilon$, for some $\epsilon \in \Gamma$. $(S, \mathrm{S}_\Gamma) \models \triangle\epsilon$ iff $\epsilon \in C$, iff $x_{\triangle\epsilon,S} \in val_S$, iff $val_S \models x_{\triangle\epsilon,S}$, iff $val \models \mathrm{tr}(\triangle\epsilon, S)$.

In the sequel, w.l.o.g. we consider that $\Box\varphi$ is an abbreviation of $[\emptyset]\varphi$. THerefore, there are 3 cases in the induction step. The cases for the Boolean operators $\neg$ and $\wedge$ are straightforward. We show the third case in the sequel.

- Let $\varphi = [E]\psi$. We start by showing that, for any two states $S, S' \in \mathrm{S}_\Gamma$, we have:
$$S \Rightarrow^E S' \text{ iff } val \models T^{E}_{(S,S')}. \tag{3}$$
There are two conditions to be checked:
(i) $C' = (C \setminus \bigcup_{p\in\mathbb{P}(E)} \mathcal{L}_{\mathrm{EQ}}(p)) \cup E$, and
(ii) $S' \in \mathrm{S}_\Gamma$ (i.e., S' is a state).
We consider three cases:
  - $\epsilon \in (C \cap \bigcup_{p\in\mathbb{P}(E)} \mathcal{L}_{\mathrm{EQ}}(p)) \setminus E$.
    It is true iff $\epsilon \notin C'$, iff $(S', \mathrm{S}_\Gamma) \not\models \triangle\epsilon$, iff $x_{\triangle\epsilon,S'} \notin val_{S'}$, iff $val_{S'} \not\models x_{\triangle\epsilon,S'}$, iff $val \models \neg x_{\triangle\epsilon,S'}$.
  - $\epsilon \in E$.
    It is true iff $\epsilon \in C'$, iff (i) $(S', \mathrm{S}_\Gamma) \models \triangle\epsilon$ and (ii) $(S', \mathrm{S}_\Gamma) \models \epsilon$, iff (i) $val \models x_{\triangle\epsilon,S'}$ and (ii) $val \models \mathrm{tr}(\epsilon, S')$ (by the induction hypothesis), iff $val \models x_{\triangle\epsilon,S'} \wedge \mathrm{tr}(\epsilon, S')$.
  - Otherwise.
    It is is true iff ($\epsilon \in C$ iff $\epsilon \in C'$), iff (i) $((S, \mathrm{S}_\Gamma) \models \triangle\epsilon$ iff $(S', \mathrm{S}_\Gamma) \models \triangle\epsilon)$ and (ii) if $(S, \mathrm{S}_\Gamma) \models \triangle\epsilon$ then $(S', \mathrm{S}_\Gamma) \models \epsilon$, iff (i) ($val \models x_{\triangle\epsilon,S}$ iff $val \models x_{\triangle\epsilon,S'}$) and (ii) if $val \models x_{\triangle\epsilon,S}$ then $val \models \mathrm{tr}(\epsilon, S')$ (by the induction hypothesis), iff $val \models (x_{\triangle\epsilon,S} \leftrightarrow x_{\triangle\epsilon,S'}) \wedge (x_{\triangle\epsilon,S} \rightarrow \mathrm{tr}(\epsilon, S'))$.

  Therefore, $S \Rightarrow^E S'$ if and only if $val \models \bigwedge_{\epsilon\in\Gamma} g(\epsilon, E, S, S')$. This ends the proof of (3).

  Now, we have $(S, \mathrm{S}_\Gamma) \models [E]\psi$ iff for all $S' \in \mathrm{S}_\Gamma$, if $S \Rightarrow^E S'$ then $(S', \mathrm{S}_\Gamma) \models \psi$, iff for all $S' \in \mathrm{S}_\Gamma$, if $S \Rightarrow^E S'$ then $val \models \mathrm{tr}(\psi, S')$ (by the induction hypothesis), iff for all $S' \in \mathrm{S}_\Gamma$, if $val \models T^{E}_{(S,S')}$ then $val \models \mathrm{tr}(\psi, S')$ (by (3)), iff $val \models \forall W_{S'} (T^{E}_{(S,S')} \rightarrow \mathrm{tr}(\psi, S'))$, iff $val \models \mathrm{tr}([E]\psi, S)$.

Now, we have: $\models_{\mathrm{QBF}} \forall W_S (D_S \rightarrow \mathrm{tr}(\varphi, S))$ iff, for all truth assignments m of the variables in W_S, if $m \models_{\mathrm{QBF}} D_S$ then $m \models_{\mathrm{QBF}} \mathrm{tr}(\varphi, S))$, iff $val_S \models \mathrm{tr}(\varphi, S)$ (because val_S is the only truth assignment of the variables in W_S that satisfies D_S), iff $val \models \mathrm{tr}(\varphi, S)$, iff $(S, \mathrm{S}_\Gamma) \models \varphi$ (as seen above). □