From Cause-in-Fact to Legally Relevant Cause through Argumentation

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

Henry Prakken
Department of Information and Computing Sciences, Utrecht University
The Netherlands
h.prakken@uu.nl

Giovanni Sartor
Law Faculty-CIRSFID, University of Bologna
European University Institute
Italy
giovanni.sartor@unibo.it

Abstract

We propose a novel framework to model legally relevant causality integrating causation-in-fact with the further conditions that are needed for it to be legally relevant. This framework is built by combining, within a modular-argumentation approach, the assessment of causation in fact and the determination of its legal relevance. For the assessment of causation-in-fact, we rely on a refinement of a previously proposed model, while for the determination of legal relevance, we propose an argumentation-based approach for defining the conditions needed for causation-in-fact to be legally relevant. We propose logical analyses of such conditions, to capture different regulatory and doctrinal approaches. Finally, we present an implementation of our framework in a logic-programming based environment for structured argumentation (Arg2P).

CCS Concepts

• **Computing methodologies → Artificial intelligence**;

Keywords

Cause-in-fact, Causation-in-law, Argumentation

ACM Reference Format:
Giuseppe Pisano, Henry Prakken, and Giovanni Sartor. 2026. From Cause-in-Fact to Legally Relevant Cause through Argumentation. In *21th International Conference on Artificial Intelligence and Law (ICAIL 2026), June 8–12, 2026 Singapore.* ACM, New York, NY, USA, 10 pages.

1 INTRODUCTION

We propose a novel, argument-based approach to model legally relevant causality, building upon the approach to cause-in-fact, also called actual causation, in [23]. Our approach is based on the assumption, extensively argued for by [28] that causal analysis in law should not substitute general models of actual causality, but rather complement them. In other words, when we move from a general model of actual causation to a model of legally relevant causation, some further aspects are to be considered, namely, those such that their presence (or absence) is needed for actual causality to be legally relevant, as a precondition for legal liability.

We assume indeed that for tortious liability all of the following must typically be present[1]:

1. the harm must actually have been caused by a wrongful act;
2. causation must have taken place in a way or context that make it relevant to the law;
3. the liable agent must be in an appropriate relation to the wrongful act (being the originator or being responsible for the person originating it);
4. the agent must have acted intentionally or negligently (except for cases of strict liability);
5. there must not exist circumstances that exclude liability (such as self defence).

The work in [23] only focused on item (1), namely, it provided an argumentation-based model of actual causation in legal cases, to which we will refer as cause-in-fact. Let us recall that the analysis of actual causation is concerned with the casual relation between singular facts (also called token-causation), namely, it is meant to determine whether a fact A is a cause of effect E (both having taken place) according to a causal theory. In legal cases, typically A is a human action and E is a harmful event. The model was inspired by the NESS (Necessary Element of a Sufficient Set) approach to causality, according to which a fact A is a cause of an effect E iff A is a necessary component of a set S of conditions which are jointly sufficient for E to be produced according to a causal theory (see [4, 13, 16, 27]). NESS overcomes the traditional idea that a cause must be a necessary condition (a *conditio sine qua non*) of the effect, an idea that does not fit cases in which multiple causal processes concur or interact. The NESS model can also cover preemption cases, through the requirement that causal laws must be instantiated in the concrete case at stake [27] (footnote 14).

These are cases in which each of two sets of conditions would be separately sufficient to produce the effect in the absence of the other, but one of such sets preempts the other from contributing to the effect. Consider, for instance, the case in which A shoots a sleeping victim and then B shoots the dead body (while ignoring that the victim is dead): each shot, in the absence of the other shot, would be sufficient to cause the victim's death, but only the first one is causally effective with regard to the death. Here we shall recover the model of [23], and refine it to better capture instances of negative causation (when the effect depends on the absence of facts which would, directly or indirectly, preempt other facts from causing the effect.

ICAIL 2026, Singapore

---

[1] For simplicity, we focus here on torts strictly understood, without considering other sources of liability, such as engaging in dangerous activities, see [12], or ownership of things or animals.

The main contribution of this paper consists in providing a framework to model item (2) above, namely, to complement causation-in-fact (actual causality) with the conditions to be met for causation-in fact to be legally relevant. We are aware that different legal approaches exist for specifying such conditions in different legal systems and also in different doctrinal approaches within the same legal systems. The very idea of a legally relevant causality is specified by using different terminologies: the legally relevant cause is referred to as a "cause-in-law", a "legal cause", a "proximate cause", an "adequate cause", a "substantial factor" of the effect at stake. Various criteria have been proposed for determining when a cause is proximate, adequate, or substantial. For instance, it has been affirmed that a cause, to be substantial, should ex ante increase the probability of the effect, should make the effect ex-ante foreseeable to a certain degree, should normally (not only exceptionally) determine the effect, or should pertain to the risks that the law intends to prevent. Moreover, it has been claimed that the causality relation is not legally relevant where it is dependent on an extraordinary and unexpected event, or on an event which is outside the risks that the law intended to address. There are different conditions to qualifying what counts as extraordinary or unexpected, or beyond the risks at stake.

Our leading idea is that the integration of cause-in-fact with these further tests can take place with the modular argument-based framework proposed in [26]. In this framework, it is assumed that argumentation can take place within different modules, possibly according to different reasoning models. Such modules can dialectically interact, where one model asks other modules to perform reasoning tasks, receiving back the results of such tasks. In the present instantiation of the modular framework, the module for computing legally relevant causation delegates the task of identifying actual causes to a causation-in-fact module. Here, the latter module could apply the argumentative model in [23], but in principle it could rely on some other model for actual causation (see [13] and [19]). After receiving this specification, the first module can then proceed to the examination of the further conditions needed from causation-in-fact to be legally relevant. As we shall see, this may involve further calls to the causation-in-fact module.

Although our main goal is to analyse and clarify legal theories of legally relevant causation, we also provide a full implementation of our approach within Arg2P [8], an environment based on logic-programming for structured argumentation.

As we cannot provide a formalisation of all different approaches to legally relevant causality, we shall use R.W. Wright's doctrinal proposal [27, 30] as our test case. According to this approach, three general conditions impede causation-in-fact from being legally relevant: no-worse-off, superseding cause, and risk playout limitation. According to Wright [30] these conditions should be viewed as defences against a claim of legal responsibility based on actual causality; we correspondingly model them as exceptions undercutting a defeasible entailment between causation-in-fact and legally relevant causality, which, in its turn, conditions legal responsibility. While extensively focusing on this interesting and coherent approach (which has inspired the most recent version of the US Restatement of the law of torts, see [30]), we shall briefly comment on how our framework can also be used to address different doctrinal approaches (based on the idea that legally relevant causation must be built on top of causation-in-fact).

The paper is structured as follows. In Section 2 we provide formal preliminaries for our work. Although this account is needed for our work to be self-contained, it will be as compact as possible (for a more extensive account, see [21]). In Section 3 we provide a model for causation-in-fact by refining the approach presented in [23]. In Section 4 we introduce our approach to move from causation in fact to legally relevant causation, through modular argumentation. In Section 5 we will apply our approach to Wright's doctrine [27, 30], making legal relevance dependent on the absence of the defences (no-worse-off, superseding cause, risk playout limitation). We shall also model cases that instantiate each of these defences. In Section 6 we shall consider how our approach may capture other doctrines on the legal relevance of causation, such as those that cover omissions or are based on notions of adequate or proximate causation. In Section 7 we present an implementation of our model in the Arg2P framework [8]. Finally, in Section 8 we discuss related work and in Section 9 we summarise our conclusions.

2 FORMAL PRELIMINARIES

An *abstract argumentation framework* [11] is a pair $AF = (\mathcal{A}_{AF}, \mathcal{D}_{AF})$, where $\mathcal{A}_{AF}$ is a set of arguments and $\mathcal{D}_{AF} \subseteq \mathcal{A}_{AF} \times \mathcal{A}_{AF}$ is a relation of attack. We write $A \in AF$ as shorthand for $A \in \mathcal{A}_{AF}$ and we will omit the subscripts if there is no danger of confusion. We will sometimes in text present an AF as $A \leftarrow B \leftrightarrow C$, to denote that $\mathcal{A} = \{A, B, C\}$ and $\mathcal{D} = \{(B, A), (B, C), (C, B)\}$.

The *ASPIC*$^+$ framework [20, 22, 25] defines abstract argumentation systems as structures consisting of a logical language $\mathcal{L}$ and two sets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules defined over $\mathcal{L}$.

In this paper, for simplicity, we assume that $\mathcal{L}$ is a set of propositional or predicate-logic literals. However, all new definitions proposed in this paper either directly apply or can be easily adapted to more expressive languages. For reasons of space we refer to [23] for examples illustrating the definitions.

Definition 2.1. [**Argumentation System**] An *argumentation system* (AS) is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a set of literals such that for every atomic formula $\varphi \in \mathcal{L}$, also $\neg\varphi \in \mathcal{L}$;
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a finite set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\{\varphi_1, \ldots, \varphi_n\} \rightarrow \varphi$ and $\{\varphi_1, \ldots, \varphi_n\} \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over literals in $\mathcal{L}$), such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$. Here, $\varphi_1, \ldots, \varphi_n$ are called the *antecedents* and φ the *consequent* of the rule.
- n is a partial function such that $n : \mathcal{R}_d \longrightarrow \mathcal{L}$.

Informally, $n(r)$ is an atomic formula (wff) in $\mathcal{L}$ which says that the defeasible rule $r \in \mathcal{R}$ is applicable, so that an argument claiming $\neg n(r)$ attacks an inference step in the argument using r. We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$. We use $\rightsquigarrow$ as a variable ranging over $\{\rightarrow, \Rightarrow\}$. Since the order of antecedents of a rule does not matter, we sometimes write $S \rightsquigarrow \varphi$ where S is the set of all antecedents of the rule.

Definition 2.2. [**Knowledge bases**] A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the *axioms*) and $\mathcal{K}_p$ (the *ordinary premises*).

Definition 2.3. [**Argumentation theories**] An *argumentation theory* is a pair $(AS, \mathcal{K})$ where AS is an argumentation system and $\mathcal{K}$ a knowledge base in AS.

Definition 2.4. [**Arguments**] An *argument A* on the basis of an argumentation theory AT is a structure obtainable by applying one or more of the following steps finitely many times:

1. φ if $\varphi \in \mathcal{K}$ with: $\mathtt{Prem}(A) = \{\varphi\}$; $\mathtt{Conc}(A) = \varphi$; $\mathtt{Sub}(A) = \{\varphi\}$; $\mathtt{Rules}(A) = \emptyset$; $\mathtt{DefRules}(A) = \emptyset$; $\mathtt{TopRule}(A) =$ undefined.
2. $A_1, \dots, A_n \rightsquigarrow \psi$ if $A_1, \dots, A_n$ are arguments such that $\psi \notin \mathtt{Conc}(\{A_1, \dots, A_n\})$ and $\mathtt{Conc}(A_1), \dots, \mathtt{Conc}(A_n) \rightsquigarrow \psi \in \mathcal{R}$ with:
   $\mathtt{Prem}(A) = \mathtt{Prem}(A_1) \cup \ldots \cup \mathtt{Prem}(A_n)$;
   $\mathtt{Conc}(A) = \psi$;
   $\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n) \cup \{A\}$;
   $\mathtt{Rules}(A) = \mathtt{Rules}(A_1) \cup \ldots \cup \mathtt{Rules}(A_n) \cup \{\mathtt{Conc}(A_1), \dots, \mathtt{Conc}(A_n) \rightsquigarrow \psi\}$;
   $\mathtt{DefRules}(A) = \mathtt{Rules}(A) \cap \mathcal{R}_d$;
   $\mathtt{TopRule}(A) = \mathtt{Conc}(A_1), \dots, \mathtt{Conc}(A_n) \rightsquigarrow \psi$.

$\mathtt{Prem}_n(A) = \mathtt{Prem}(A) \cap \mathcal{K}_n$ and $\mathtt{Prem}_p(A) = \mathtt{Prem}(A) \cap \mathcal{K}_p$. Furthermore, argument A is *strict* if $\mathtt{DefRules}(A) = \emptyset$ and *defeasible* otherwise, and A is *firm* if $\mathtt{Prem}_p(A) = \emptyset$, otherwise A is *plausible*.

The set of all arguments on the basis of AT is denoted by $\mathcal{A}_{AT}$.

Each function $\mathtt{Func}$ in this definition is also defined on sets of arguments $S = \{A_1, \dots, A_n\}$ as follows: $\mathtt{Func}(S) = \mathtt{Func}(A_1) \cup \ldots \cup \mathtt{Func}(A_n)$. Note that the $\rightarrow$ and $\Rightarrow$ symbols are overloaded to denote both inference rules and arguments.

Definition 2.5. [**Attack**] Argument A *attacks* argument B iff A *undercuts* or *rebuts* or *undermines* B, where:

- *A undercuts B* (on B') iff $\mathtt{Conc}(A) = -n(r)$ and $B' \in \mathtt{Sub}(B)$ such that B''s top rule r is defeasible.
- *A rebuts B* (on B') iff $\mathtt{Conc}(A) = -\varphi$ for some $B' \in \mathtt{Sub}(B)$ of the form $B''_1, \dots, B''_n \Rightarrow \varphi$.
- *A undermines B* (on φ) iff $\mathtt{Conc}(A) = -\varphi$ for some $\varphi \in \mathtt{Prem}(B) \cap \mathcal{K}_p$.

Abstract argumentation frameworks are then generated from an argumentation theory AT as follows:

Definition 2.6. [**Argumentation frameworks**] An *abstract argumentation framework (AF)* defined by an argumentation theory AT is a pair $(\mathcal{A}, C)$ where $\mathcal{A}$ is the set of all arguments on the basis of AT and $(X, Y) \in C$ iff X attacks Y.

Here we use labelling-style semantics [3, 11] for evaluating arguments within argumentation frameworks. A *labelling* of a set of arguments in an $AF = (\mathcal{A}, \mathcal{D})$ is any triple of non-overlapping subsets (*in*,*out*,*und*) of $\mathcal{A}$ that satisfies the following constraints:

1. an argument is *in* iff all arguments defeating it are *out*;
2. an argument is *out* iff it is defeated by an argument that is *in*;
3. an argument is *und* iff it is neither *in* nor *out*.

The *grounded labelling* of an AF minimises the set of arguments that are labelled *in* and is always unique. A set $S \subseteq \mathcal{A}$ is called the *grounded extension* of AF iff S is the set of all arguments labelled *in* in the grounded labelling.

Lastly, we introduce the notion of *grounded argument games* [24], where a proponent and an opponent of an argument A alternatively exchange arguments, with each move consisting on an argument from AF. The proponent starts with A and then the opponent has to defeat the last argument of the proponent while the proponent has to asymmetrically defeat the last argument of the opponent. Moreover, the proponent is not allowed to repeat its own moves. A game is won if the other player has no replies. So a player has a winning strategy if they can make the other player run out of moves.

3 CAUSALITY IN FACT

In this section, we summarise the model of cause-in-fact presented in [23] and propose a refinement of it.

Definition 3.1. [**Explanation**] An *explanation* for an argument A relative to argumentation framework AF is the set of all proponent arguments in any minimal winning strategy for A in the grounded argument game relative to AF

Note that an explanation, as a minimal winning strategy for A, identifies a minimal set of arguments E which (a) contains A and (b) can overcome all objection against any argument in E.

We now introduce the notions of *causal theory* and its *axiom-expansion.*

Definition 3.2. [**Causal Theory**] A *causal theory CT* is an argumentation theory $(AS, \mathcal{K}_n \cup \mathcal{K}_p)$ such that $\mathcal{K}_n = \emptyset$.

The axiom-expansion consists in adding new indefeasible axioms to a causal theory. The added axioms override any incompatible defeasible premise or conclusion, providing a mechanism for counterfactual reasoning: the expanded knowledge base provides the conclusions that would be obtained if the axioms were true (rather than the incompatible defeasible conclusions derivable from the original knowledge base). The axiom-expansion of a causal theory with a set of literals L is similar to an intervention in [15].

Definition 3.3. [**Axiom-expansion**] The *axiom expansion of a causal theory* $CT = (AS, \mathcal{K})$ with intervention L is defined as $CT + L = (AS, \mathcal{K} \cup L)$, where $L \subseteq \mathcal{L}$ is a set of axioms.

We now propose a revision of the definition of *Cause* in [23] to provide a more intuitive account (directly linking causes and explanations).

Definition 3.4. [**NESS Cause**] Literal φ is a *cause* of literal ψ relative to argumentation framework AF_{CT} iff there exist

- an argument $D \in \mathcal{G}(AF_{CT})$ with $\mathtt{Conc}(D) = \varphi$,
- an argument A with $\mathtt{Conc}(A) = \psi$,
- an explanation E for A,
- a set of literals $L \subseteq \mathcal{L}$ with $-\phi \in L$,
- an argument $B \in AF_{CT+L}$ with $L \subseteq Sub(B)$

such that

- B attacks an argument $C \in E$
- B labelled either *in* or *und* in $\mathcal{G}(AF_{CT+L})$

According to this definition, for a proposition φ to be a cause of ψ, there must be an explanation E for an argument A for φ that no longer holds in the argumentation framework expanded with L (which includes $-\varphi$). For this to be the case, L must enable the construction of an effective counterargument B against (an argument C in) E, namely a counterargument that is not *out* (being *in* or *und*). This would happen in two cases: *(1)* where an argument in E contains (a subargument with) conclusion φ, which would be defeated by the argument $\{-\varphi\}$; *(2)* where $-\varphi$, possibly with the addition of further literals (included in L), enables the construction of a new argument that attacks an argument in E.

Here is a simple example of the two cases. Consider argumentation theory $AT = (\mathcal{R}, \mathcal{K})$:

$$\mathcal{R} = \quad \{r_0 : a \Rightarrow b, \quad r_1 : c \Rightarrow \neg r_0\} \quad \mathcal{K} = \{a, \neg c\}$$

In the argumentation framework AF_{AT} we have an explanation $E = \{A\}$ of b, where $A = \{a\} \Rightarrow_{r_0} b$. There are two NESS causes for b: *(i)* a since in since in $AF_{AT+\{\neg a\}}$, we have an *in*counterargument to A, namely $\{\neg a\}$; *(ii)* $\neg c$ since in $AF_{AT+\{c\}}$ we also obtain an *in* counterargument to A, namely $\{c\} \Rightarrow_{r_1} \neg r_0$.

We introduce a definition of a *but-for cause*, i.e., of causes that are necessary for the effect (can be generalized to sets of literals).

Definition 3.5. [**But-for Cause**] Literal φ is a *but-for cause* of literal ψ relative to AF_{CT} iff there exist: (a) an argument A with $Conc(A) = \psi$ in $\mathcal{G}(AF_{CT})$ and (b) no such argument in $\mathcal{G}(AF_{CT+\{-\varphi\}})$.

Clearly, a NESS cause φ may not be a but-for cause in cases of overdetermination, as shown in the following example from [23].

Asbestos and Lung Cancer (Overdetermination): Fairchild v Glenhaven (2002). [2]

Mr. Fairchild had worked for three construction companies during the 1960s, where he frequently handled asbestos materials. He passed away from lung cancer in 1996. Expert testimony confirmed that asbestos exposure from at least two of these companies was enough to cause his cancer. Consequently, all three companies were found liable for his death.

The atoms of the scenario: Fairchild was exposed to asbestos in i (Ex_i); Fairchild contracted lung cancer (Ca); Fairchild died (Di)[3].

$$\begin{aligned} \mathcal{R} = \{ \quad & r_0 : Ca \Rightarrow Di, && r_1 : Ex_1, Ex_2 \Rightarrow Ca, \\ & r_2 : Ex_2, Ex_3 \Rightarrow Ca, && r_3 : Ex_1, Ex_3 \Rightarrow Ca\} \\ \mathcal{K} = \{ \quad & Ex_1, Ex_2, Ex_3\} \end{aligned}$$

Each Ex_i (with $i = 1, 2, 3$), i.e., the exposure while working in company i, is a NESS cause (Definition 3.4) of Di, since each Ex_i is included in an explanation for Di. In contrast, each Ex_i (with $i = 1, 2, 3$) is not a But-for cause (Definition 3.5) of Di, since the lack of a single exposure would not be sufficient to prevent Fairchild's death in the framework.

4 LEGALLY RELEVANT CAUSES

As noted earlier, there is an important distinction between cause-in-fact and legally relevant causes: in some instances cause-in-fact fails to provide an adequate ground for legal liability. Following [27, 30], we think that cause-in-fact should be a key necessary element for legal responsibility, though it may need to be integrated with further aspects. We will we speak of legally relevant causation to include causation-in-fact plus the further aspects that are needed for causation in fact to determine legally relevant causality.

In different legal systems (and according to different doctrines and cases within the same legal systems), the conditions of cause-in-fact to be legally relevant have been specified in different ways. Some approaches require a degree of "foreseeability" as a necessary condition in addition to causality-in-fact. Other approaches require that the cause-in-fact would ex ante determine an increase in the probability of event. However, exceptions to such general principles can be found. Consider, for instance, the principle that a defendant must "take the victim as they find them", i.e., be liable for the harm resulting from a plaintiff's unforeseeable and uncommon reactions. Accordingly, if the victim has a very rare health condition which makes them exceptionally susceptible to harm (e.g. hemophilia), causation-in-fact may engender liability even in the absence of ex ante foreseeability (the hemophilic victim dies as a consequence of a punch that would normally only have caused a minor injury).

Other approaches, rather than defining additional requirements for actual causality to be legally relevant, focus on the conditions which prevent causation-in-fact from being legally relevant, treating such conditions as possible defences. For instance, an actual cause may not be legally relevant when exceptional events or circumstances decisively contribute to determine an unexpected causal outcome. Such an exclusion may in its turn be subject to exceptions, for instance, according to the above mentioned rule that defendant must "take the victim as they find them".

We believe that for capturing all these different approaches to legally relevant causality, the argumentation approach is very suitable, as it allows to express the general principle for establishing or excluding the legal relevance of causes-in-fact, as well as exceptions to these principles (and possible exceptions to exceptions).

To integrate causation-in-fact with such additional conditions and/or exceptions, we will rely on the modular approach to argumentation developed in [26]. The key idea of this approach is that a reasoning problem can be decomposed by a metalevel problem specification that connects a set of problem solving modules via output-input rules. Each module receives input from and can provide input to other modules. A problem specification is itself an *ASPIC*$^+$ theory, in which the logical language $\mathcal{L}$ combines the metalanguages of the individual problem solving modules and in which the output-input rules are inference rules over this metalanguage that have as antecedents the outputs of one or more modules and have as consequent a single input to another module. The framework of [26] is liberal with regard to the reasoning to be deployed in each module: it may consist in defeasible argumentation, but also, for instance, in deductive reasoning or Bayesian inference.

For our purposes, we do not need to recall all the definitions of [26]; it is sufficient to show how it can be applied in our case. What we need are two problem solving modules. The first is a causal argumentation module *CIF* of the form specified in Section 3, whose input is the causal theory CT of a legal case, and whose output is that a certain fact is a cause-in-fact of a certain effect, relative to that input theory. The second module we need is an *ASPIC*$^+$-style legally-relevant-causation module *LRC*, which can be used to determine which causes-in-fact are legally relevant. The

---

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

[3] In all our examples below the antecedents of causal rules that cannot be derived from other causal rules are in K_p.

metalevel problem specification connects these two modules with the following output-input rule, which transforms outputs of a cause-in-fact module *CIF* to inputs of the legally-relevant-causation module *LRC* (we leave the variables in the oi rule name implicit):

$$\begin{aligned} oi(\ldots) : \varphi \text{ is a NESS cause of } \psi \text{ relative to } AF_{CT} \Rightarrow \\ CaFa(\varphi, \psi) \in \mathcal{K}_n(LRC). \end{aligned} \tag{1}$$

Informally this says that if it is justified in the cause-in-fact module *CIF* that φ is a NESS cause of ψ, then the fact $CaFa(\varphi, \psi)$ is added as an axiom in the legally-relevant-causation module *LRC*. For an implementation of the *CIF* and *LRC* modules see Section 7.

5 WRIGHT'S APPROACH FORMALISED

To capture Wright's approach, we first define a general defeasible axiom according to which every cause-in-fact is in principle legally relevant, and then consider what exceptions may apply to this axiom.

As specified in the previous section, we assume that the language of the legally-relevant-causation module *LRC* contains literals of the form $CaFa(x, y)$. The $ASPIC^+$-style *LRC* module then internally specifies with a defeasible rule that if φ is a cause-in-fact of ψ relative to the causal module *CIF* (for simplicity we omit references to *CIF*) then presumably φ is a legally relevant cause of ψ:

$$CF2LRC(\varphi, \psi) \ : \ CaFa(\varphi, \psi) \Rightarrow LeReCa(\varphi, \psi) \tag{2}$$

Wright distinguishes three main exceptions (defences).

No-Worse-Off. According to the no-worse-off exception, causation-in-fact is not legally relevant if the harm would have occurred anyway in the absence of any tortious conduct or condition. For example, in *Kingston v. Chicago & Northwestern Railway Co. (191 Wis. 610, 211 N.W. 913 1927)* [29] the defendant negligently set fire but another fire of unknown origin, each of which was independently sufficient to destroy the plaintiff's property, merged together and destroyed the property. The Supreme Court of Wisconsin stated that the defendant would not be liable for the destruction of the plaintiff's property if it proved that the other fire had a natural (and thus nontortious) origin.

The rule for the exception states that if there is a parallel non-tortious cause causing the effect, then rule *CF2LRC* does not apply.

$$\begin{aligned} NWO(\varphi, \psi, \chi) : CaFa(\chi, \psi),\ \chi \neq \varphi,\ \neg Tortious(\chi) \\ \Rightarrow \neg CF2LRC(\varphi, \psi) \end{aligned} \tag{3}$$

Atoms in *Kingston v. Chicago & Northwestern Railway Co.* are: defendant sets fire ($DeFi$); second unknown fire ($UnFi$); property ignites (Ig); property destroyed (De).

$$\begin{aligned} \mathcal{R} = \{ \quad & r_0 : Ig \Rightarrow De, \quad r_1 : DeFi \Rightarrow Ig, \quad r_2 : UnFi \Rightarrow Ig\} \\ \mathcal{K} = \{ \quad & DeFi, UnFi\} \end{aligned}$$

This is a case of overdetermination, and Definition 3.4 correctly identifies both $DeFi$ and $UnFi$ as causes-in-fact of De. Let us now assume a second framework containing both schemes *CF2LRC* and NWO, and the following knowledge base resulting from the cause-in-fact evaluation:

$$\begin{aligned} \mathcal{K} = \{ \quad & CaFa(DeFi, De),\ CaFa(UnFi, De), \\ & \neg Tortious(UnFi),\ DeFi \neq UnFi\ \} \end{aligned}$$

Instantiating the theory yields seven arguments: one for each fact in the knowledge base, one deriving $LeReCa(DeFi, De)$ via *CF2LRC*, one deriving $LeReCa(UnFi, De)$ via *CF2LRC*, and one deriving $\neg CF2LRC(DeFi, De)$ via the NWO exception. The grounded extension therefore concludes that the defendant's behaviour is not a legally relevant cause of the damages to the plaintiff's property, since the argument for $LeReCa(DeFi, De)$ is undercut by the presence of the parallel non-tortious cause $UnFi$.

Superseding cause. According to the superseding cause exception (which is also referred to with the term "actus novus interveniens"), causation-in-fact is not relevant to legal liability if a different cause of the plaintiff's injury *(1)* intervened between the defendant's tortious conduct and the plaintiff's injury, *(2)* was a necessary (but-for) cause of the plaintiff's injury, and *(3)* was highly unexpected. For instance in the case of *Watson v. Kentucky & Indiana Bridge & Railroad Co. (137 Ky. 619, 126 S.W. 146 1910)* [29] the defendant railroad's tank car full of gasoline derailed and spilled gasoline onto the surrounding streets and into the gutters, causing large quantities of gasoline vapors to accumulate in the air. Three hours later the vapors were lit by a match dropped or thrown by Duerr, which caused a violent explosion injuring the plaintiff. The court held that if Duerr's action was deliberate and malicious, it would be "so unexpected and extraordinary" that it would be a superseding cause that would relieve the defendant of liability for plaintiff's harm.

The rule for superseding causes says that if there is a superseding cause then rule *CF2LRC* does not apply.

$$SC(\varphi, \psi, \chi) : SupersedingCause(\chi, \varphi, \psi) \Rightarrow \neg CF2LRC(\varphi, \psi) \tag{4}$$

A superseding cause is characterised as a cause which happened after the tortious conduct, was necessary and highly unexpected.

$$\begin{aligned} SuCa(\varphi, \psi, \chi) : CaFa(\varphi, \psi),\ CaFa(\chi, \psi),\ \chi \neq \varphi,\ After(\chi, \varphi), \\ Unexpected(\chi),\ ButFor(\chi, \psi) \\ \Rightarrow SupersedingCause(\chi, \varphi, \psi) \end{aligned} \tag{5}$$

Atoms in *Watson v. Kentucky & Indiana Bridge & Railroad Co.* are: defendant spills gasoline (Ga); Duerr drops match (Ma); gasoline vapours ignite (Ig).

$$\mathcal{R} = \{r_0 : Ga, Ma \Rightarrow Ig\} \quad \mathcal{K} = \{Ga, Ma\}$$

In this case, Definition 3.4 identifies both Ga and Ma as causes of Ig (as does Definition 3.5). Let us now assume a second framework containing the schemes *CF2LRC*, *SC*, and *SuCa*, and the following knowledge base resulting from the cause-in-fact evaluation, together with the additional information that the match drop was highly unexpected and occurred after the gasoline spill ($Unexpected(Ma)$ and $After(Ma, Ga)$):

$$\begin{aligned} \mathcal{K} = \{ \quad & CaFa(Ga, In),\ CaFa(Ma, In), ButFor(Ga, In),\ Ma \neq Ga \\ & ButFor(Ma, In),\ Unexpected(Ma),\ After(Ma, Ga)\ \} \end{aligned}$$

Instantiating the theory yields eleven arguments: one for each fact in the knowledge base, one deriving $LeReCa(Ga, In)$ via *CF2LRC*, one deriving $LeReCa(Ma, In)$ via *CF2LRC*, and one deriving $\neg CF2LRC(Ga, In)$ via the *SC* exception (and thus *SuCa*). The grounded extension therefore concludes that the railroad is not legally responsible for the plaintiff's injury, as the argument for

$LeReCa(Ga, In)$ is undercut by the presence of the highly unexpected superseding cause Ma.

For another example, consider the case of *R v Jordan (40 Cr App R 152. 1956)).* The case involves a stabbing victim who was brought to a hospital, where medical negligence caused his death: while the stab wounds were healing well, the administration of an antibiotic to which the victim was intolerant caused his death. The administration of the antibiotic was considered an *actus novus*, so that Jordan escaped liability. Given premises $CaFa(JoSt, ViDe)$ (the stabbing by Jordan is a cause-in-fact of the victim's death), $Unexpected(AdAn)$ and $After(AdAn, JoSt)$ (the administration of the antibiotic is an *actus novus* between the stabbing and the death), we can conclude, via $SuCa$ and SC that $\neg CF2LRC(JoSt, ViDe)$.

Risk Playout. According to the risk-playout limitation, causation-in-fact is not legally relevant (for the purpose of liability) if the harm did not occur as part of the realization and playing out of one of the foreseeable risks that made the person's conduct tortious, before the hazards created by the realization of that risk had dissipated.

For instance, in the case of *Berry v. Sugar Notch Borough (191 Pa. 345, 43 A. 240 1899)* [30], the plaintiff was driving faster than eight miles per hour in violation of a local ordinance when a tree fell onto his car, causing injury. The defendant argued that the plaintiff's unlawful speed brought him to the precise place at the exact moment the tree fell. However, the court held that the plaintiff's speeding was not the legally relevant cause of the injury, but merely a coincidence. Under the risk-playout limitation, the plaintiff's speeding was not legally relevant to causation, because the harm did not result from the realization and playing out of any risk created by the speeding violation.

We capture the no risk-playout exception by making risk-playout a limitation to a defence based on non-foreseeability. In other terms, we assume that in general the causation of a non-foreseeable event is not legally relevant, unless the unforeseable event takes place within the playout of a foreseeable risk. Here is the general limitation of legal relevance based on non-foreseeability.

$$NoFor(\varphi, \psi) : \neg Forseeable(\varphi, \psi) \Rightarrow \neg CF2LRC(\varphi, \psi) \tag{6}$$

This exception is undercut in the case that the non-foreseeable effect ψ is within the playout of a risk χ which is a foreseeable implication of the cause φ.

$$\begin{aligned} RPL(\varphi, \psi) : ForeseeableRisk(\varphi, \chi), Playout(\chi, \psi)) \Rightarrow \\ \neg NoFor(\varphi, \psi) \end{aligned} \tag{7}$$

The predicate $Playout$, which may involve complex jurisprudential considerations (see [30]), may be defined in a further module.

Atoms in *Berry v. Sugar Notch Borough* are: Plaintiff is proceeding (*PlPr*); Plaintiff is speeding in violation of the ordinance (*PlSp*); Plaintiff is on the borough road (*PlRo*); A tree falls onto Plaintiff's car ($TrFa$); Plaintiff suffers injury (*In*).

$$\begin{aligned} \mathcal{R} = \{ \quad & r_0 : PlPr, PlSp \Rightarrow PlRo, \\ & r_1 : TrFa, PlRo \Rightarrow In \ \} \\ \mathcal{K} = \{ \quad & PlPr, PlSp, TrFa\} \end{aligned}$$

Definition 3.4 gives us that both Plaintiff's speeding (*PlSp*) and the falling tree event ($TrFa$) are causes of the injury (*In*). Let us now consider the framework containing $NoFor$, $CF2LRC$, RPL and the following set of premises:

$$\begin{aligned} \mathcal{K} = \{ \quad & CaFa(PlSp, In), \ CaFa(TrFa, In), \\ & ForeseeableRisk(PlSp, SpCo), \neg Forseeable(PlSp, In) \\ & \neg Playout(SpCo, In) \ \} \end{aligned}$$

In this case, it was not foreseeable that, by exceeding the speed limit, Plaintiff would suffer harm from a falling tree ($\neg Forseeable(PlSp, In)$). There are foreseeable risks associated with speeding, such as loss of control or inability to stop in time ($ForeseeableRisk(PlSp, SpCo)$), but being struck by a falling tree is not related to the realization of any such risk ($\neg Playout(SpCo, In)$).

When the theory is evaluated, $TrFa$ (the falling tree event) emerges as the only legally relevant cause of *In*, because Plaintiff's speeding does not satisfy the risk-playout condition. In other words, under the *RPL* scheme, Plaintiff's excessive speed is undercut as a proximate cause, leaving the falling tree event as the sole legally relevant cause of the injury.

6 Other issues on the legal relevance of causation

The three defences identified by [27, 30] do not exhaust all ways in which the legal relevance of a cause can be challenged.

Causation through omission. An interesting case concerns omissions, as in most cases, legal systems put stricter conditions for responsibility through omissive acts. Usually, agents are liable for the consequences of their omissions only where they were under a legal obligation to take action to prevent such consequences. This restriction can be modelled in different ways: as a requirement for the wrongfulness of the omission (thus outside of the causality analysis) or as a requirement for causation through omission to be legally relevant. The latter seems to be the approach adopted in the Italian criminal code, which says: "not preventing an event, the prevention of which is a legally obligation, amounts to causing". This clause is understood (reasoning "a contrario") as entailing that causation-in-fact through omissions is not legally relevant if there is no obligation to prevent the adverse effect at stake.

One way to model this limitation on the legal relevance of causality could consist in adding an additional defence. In our framework, this may be done through the omission scheme $UnOm$, which undercuts the CF2LRC scheme if the cause-in-fact at stake is an omission whose complement is not obligatory.

$$UnOm(\varphi, \psi) : Om(\varphi), \neg Obl(-\varphi) \Rightarrow \neg CFLRC(\varphi, \psi) \tag{8}$$

In our approach the issue of whether the omission is obligatory can be modelled in another problem solving module, which needs to consider that legal systems may regulate in different ways obligations to prevent harm. In general, common law systems tend to exclude an obligation to rescue when life is in danger, while civil law systems tend to recognise such an obligation. For instance, in the case *Yania v. Bigan (Penn. 1959),* Miner Yania was acquitted for failing to help the colleague Bigan who had fallen into a water pit. Here are the atoms: $YaHe$ (Yania helps Bigan), $BiDi$ (Bigan dies). Assume that our cause-in-fact module has stated that Yania by not helping Bigan caused-in-fact the latter's death: $CaFa(\neg YaHe, BiDi)$. Also assume that this is an omission ($Om(\neg YaHe)$) and that the omitted action is not obligatory under US law $\neg Obl(YaHe)$. By instantiating the just introduced scheme, we obtain the following argumentation theory:

$$\begin{aligned}
\mathcal{R} = \quad & \{CFLRC(\neg YaHe, BiDi) : CaFa(\neg YaHe, BiDi) \\
& \quad \Rightarrow LeReCa(\neg YaHe, BiDi), \\
& UnOm(\neg YaHe, Bidi) : Om(\neg YaHe), \neg Obl(YaHe) \\
& \quad \Rightarrow \neg CFLRC(\neg YaHe, BiDi)\} \\
\mathcal{K} = \quad & \{CaFa(\neg YaHe, BiDi), Om(\neg YaHe), \neg Obl(YaHe)\}
\end{aligned}$$

We can build a justified argument for $\neg CFLRC(\neg YaHe, BiDi)$, excluding that the causation at stake is legally relevant. Note that the omission exception in this case would not have been satisfied in legal systems in which (contrary to the US system) there is an obligation to help, when death or serious harm are at stake.

For an example in which, on the contrary, the omission exception was not satisfied (the omitted action being legally obligatory), consider an Italian case (Case 12124/2023, Cassazione Penale) in which parents were condemned for not providing leukemia treatment to their child. In this case, the judges considered that parents were obliged to care for the child, providing the available medical treatment (for a logical analysis see [23]).

Adequate/promimate causality. Our idea of combining, within a modular argumentation approach, causation-in-fact may also be used to model legal doctrines in which the causation-in-fact is complemented by further positive conditions rather than by defences. For instance, in Italian criminal law the approach denoted “adequate causality” (causalità adeguata), according to which a casual connection is only relevant when the effect is a normal or at least a probable consequence of the action, is often adopted. This idea expresses the requirement of ex ante foreseeability, as the concept of proximate causality in common law jurisdictions.

This approach could be modelled by substituting rule $CF2LRC$ with the following

$$\begin{aligned}
CF2LRCBis(\varphi, \psi) \ : & \ CaFa(\varphi, \psi), \ CausesNormally(\varphi, \psi) \\
& \Rightarrow LeReCa(\varphi, \psi)
\end{aligned} \tag{9}$$

where $CausesNormally(\varphi, \psi)$ is a predicate referring to a general probabilistic connection (based on laws connecting kinds of events). The specification of whether this connection holds may be delegated to a further reasoning module for (defeasible) type-causality.

Causality and increase in probability. Similarly, our framework could be used to capture the idea advanced by certain law-and-economics authors, according to which causality is legally relevant only if the cause ex ante increases rather than reduces the probability of the effect [17]. Consider, for instance, the case in which an object is sent by rail rather than by road, and a train accident destroys the object. In this case, sending the object by rail is an actual cause of the harm to the owner of the object (the object would have safely arrived had it been sent by road). However, liability may be excluded as, ex ante, rail was safer than road. This approach can be implemented through a rule as the one used to capture normal (adequate) causation, just by substituting the predicate $CausesNormally$ with the predicate $IncreasesProbability$.

Ennoblement. According to some views (for a discussion see [4]), causality should be excluded when each of φ and $-\varphi$ is included in a (different) causal process leading to ψ, so that both φ and $-\varphi$ equally contribute to ψ. For instance, consider the case in which the cruel sergeant tells the soldier: “Shoot the prisoner! If you do not, I will do it myself (in [23] such cases are referred to with the term ennoblement [31]). Within our framework the soldier would equally cause the prisoner to die by shooting and by not doing so. We may address such contexts by restricting the notion of causality with an additional condition (requiring that it is not the case that both ϕ and $\neg\phi$ determine the same effect). More consistently with the NESS idea, we prefer to accept that there is causation-in-fact, but argue that it is not legally relevant (so that there is no responsibility). Such a limitation of responsibility is especially intuitive where the harmful outcome is less probable according the alternative not implemented. Thus, we might argue that the soldier would be liable if he shoots, while he would not be liable if he does not shoot, since there is tenuous chance that in the latter case the sergeant will not implement her threat.

7 IMPLEMENTATION

In this section, we present an implementation of the system described in Section 4, combining a cause-in-fact solver with a legally relevant causation module within the Arg2P framework [8]. The implementation closely follows the outlined modular architecture, where causation-in-fact is computed independently, and its results are then filtered through a defeasible argumentation layer that determines legal relevance.

The system consists of two main components:

1. A *cause-in-fact module*, which evaluates whether a fact φ is a cause-in-fact of an effect ψ relative to a causal theory.
2. A *legally relevant causation module*, which determines whether a cause-in-fact should be considered legally relevant.

The modules are loosely coupled: the cause-in-fact solver is queried as an external reasoning component, while the legally relevant causation module reasons defeasibly over its outputs. A *meta-level bridge* connects the output of the cause-in-fact module to the input of the legally relevant causation module via output–input rules.

7.1 CAUSE-IN-FACT MODULE

The *CIF* module implements the NESS-style test from Definition 3.4. Algorithm 1 gives high-level pseudocode for the cause-in-fact procedure. This algorithm is implemented in Arg2P by: *(i)* computing grounded argument graphs for the effect, *(ii)* extracting all supporting arguments, *(iii)* testing whether argument expansions (interventions) render these arguments *out* or *undecided*, and *(iv)* enforcing minimality via subset checks.

As an example, let us use the *Kingston v. Chicago & Northwestern Railway Co.* case. In Arg2P, causal rules and factual knowledge are encoded as follows:

Listing 1: No-Worse-Off Exception in the CIF Module

```
% R
r0 : ig => de.
r1 : deFi => ig.
r2 : unFi => ig.

% K
f1 :=> deFi.
```

Algorithm 1 NESS Evaluation

```
Require: Causal theory $T$, candidate cause $\varphi$, effect $\psi$
Ensure: $\varphi$ satisfies the NESS condition for $\psi$
Compute grounded labelling of $T$
Extract supporting arguments of $\psi$ from $T$
if no supporting arguments exist then return false
end if
Generate possible interventions on $T$ including $-\varphi$
for all interventions $I$ do
    Compute grounded labelling of $T + I$
    if $T + I$ makes a supporting argument out or und then
        if $I$ is minimal w.r.t. the in attacker then return true
        end if
    end if
end for
return false
```

```
f2 :=> unFi.
```

This module can be queried with the predicate `causality::ness(X, Cause, Effect)` to determine whether a given cause (`Cause`) is a cause-in-fact of a given effect (`Effect`). The intervention variable `X` is left unbound and assigned by the solver during evaluation. Given the theory in Listing 1, both queries `causality::ness(X, deFi, de)` and `causality::ness(X, unFi, de)` evaluate to `true`, with the solver assigning `X = -deFi` and `X = -unFi`, respectively.

The module also implements the *But-For test* from Definition 3.5. The test verifies whether the effect ψ necessarily depends on the candidate cause φ. Algorithm 2 gives high-level pseudocode for the procedure. We conclude by illustrating the But-For test on the example in Listing 1. To this end, the *CIF* module provides the predicate `causality::but_for(Cause, Effect)`, which succeeds whenever the candidate cause satisfies the But-For condition with respect to the given effect.

Algorithm 2 But-For Evaluation

```
Require: Causal theory $T$, candidate cause $\varphi$, effect $\psi$
Ensure: $\varphi$ satisfies the But-For condition for $\psi$
Compute grounded labelling of $T$
if no argument for $\psi$ is labelled in then return false
end if
Compute grounded labelling of $T + -\varphi$
if no argument for $\psi$ is labelled in then return true
end if
return false
```

In this case, the effect `de` is supported by an *in*-labelled argument derived from either `deFi` or `unFi`. However, intervening on `deFi` does not eliminate all supporting arguments for `de`, since `unFi` still provides an independent support, and vice versa. As a consequence, neither `deFi` nor `unFi` satisfies the But-For condition for `de`. Accordingly, both queries `causality::but_for(deFi, de)` and `causality::but_for(unFi, de)` evaluate to `false`, reflecting the absence of counterfactual dependence.

7.2 LEGALLY-RELEVANT-CAUSATION MODULE

The *LRC* module is simpler than the cause-in-fact module, since its rules and exceptions can be expressed directly in a standard *ASPIC*$^+$-style theory. As an illustration, Listing 2 shows the formalization in Arg2P of the No-Worse-Off exception for the *Kingston* case.

Listing 2: No-Worse-Off Exception in the LRC Module

```
% General scheme
cf2lrc(Y, Z) : caFa(Y, Z) => leReCa(Y, Z).

% No-Worse-Off exception
nwo(Y, Z, X) : caFa(X, Z), caFa(Y, Z), prolog(X \= Y), -tortious(X) => -cf2lrc(Y
      , Z).

% Non-tortious fact: the unkonown fire is not tortious
f0 :=> -tortious(unFi).

% Input from the cause-in-fact module
f1 :=> caFa(deFi, de).
f2 :=> caFa(unFi, de).
```

When evaluated, the grounded extension of this module correctly undercuts the argument for $leReCa(deFi, de)$, reflecting that the defendant has not caused the property destruction in a legally relevant way due to the presence of the parallel non-tortious cause.

7.3 BRIDGING THE MODULES

The output of the cause-in-fact module is fed into the legally relevant causation module via a *meta-level bridge*. This bridge is formalized as the output–input rule $oi(\ldots)$ from Section 4, which transforms statements about causes-in-fact into input facts for the legally-relevant module. In Arg2P, we leverage the `prolog(...)` syntax to trigger the evaluation of cause-in-fact by the dedicated solver, ensuring that only verified causes-in-fact are passed to the *LRC* module. The bridge can be expressed as in Listing 3.

Listing 3: Meta-Level Bridge Rule Connecting Modules

```
% Meta-level bridge: transform verified causes-in-fact into inputs for LRC
oi(Y, Z) : check(Y, Z), prolog(causality::ness(X, Y, Z)) -> caFa(Y, Z).
```

The predicate `check(Y, Z)` is used to explicitly control the set of cause–effect pairs that the user intends to verify. These check predicates are user-provided input facts that must be explicitly included in the theory to trigger the evaluation process. For each such pair, the bridge rule triggers the evaluation of causation-in-fact via the dedicated solver using the `prolog(...)` construct. If the cause-in-fact verification succeeds, the term `caFa(Y, Z)` is derived as the conclusion of the argument based on check and the bridge rule `oi`, and is then made available for use by the rules of the legally-relevant-causation module.

Combining the cause-in-fact module, the legally-relevant-causation module, and the bridge, the complete theory for the *Kingston* case is shown in Listing 4.

Listing 4: Integrated Cause-in-Fact and LRC Theory

```
%----------------------------
% Cause-in-Fact Module
%----------------------------
r0 : ig => de.
r1 : deFi => ig.
r2 : unFi => ig.

f1 :=> deFi.
f2 :=> unFi.

%----------------------------
% Legally Relevant Causation Module
%----------------------------
```

```
cf2lrc(Y, Z) : caFa(Y, Z) => leReCa(Y, Z).
nwo(Y, Z, X) : caFa(X, Z), caFa(Y, Z), prolog(X \= Y), -tortious(X) => -cf2lrc(Y
       , Z).

f :=> -tortious(unFi).

%----------------------------
% Bridge Module
%----------------------------
oi(Y, Z) : check(Y, Z), prolog(causality::ness(X, Y, Z)) -> caFa(Y, Z).

c1 :-> check(deFi, de).
c2 :-> check(unFi, de).
```

This integrated theory can be executed using the Arg2P online environment[4] as shown in Figure 1.

8 RELATED WORK

The standard approach to causality in AI is the Halpern-Pearl account in terms of counterfactual reasoning with structural causal models [13]. As we explained in more detail in [23], the reason we deviate it since their account essentially regards causal laws as deductive, while we want to capture their defeasible nature.

To the best of our knowledge, there is little work in AI & law addressing the question when cause-in-fact is legally relevant. An exception is Andreas et al. [1], who model overdetermination within a counterfactual approach to causation. Their approach is not explicitly related to NESS but to the Halpern-Pearl model of actual causality [13]. It applies possible-worlds semantics and relies on the idea of 'normatively ideal worlds', which are worlds in which agents act according to their legal duties. They use this concept to model a notion of causal responsibility. Thus Andreas et al. do, unlike us, not separate cause-in-fact from legal responsibility. Another difference with our approach is that we take an argumentation approach while they take a modal-logic approach. Other AI & law work addresses different aspects of causality. Bex [7] models the use of causal relations to explain evidence in legal evidential reasoning. Lehmann & Gangemi [18] propose a formal ontology of concepts related to cause-in-fact, but do not model causal reasoning. Araszkiewicz [2] offers a semi-formal model of adequate causality within statutory interpretation. Similarly, we support a jurisdiction-sensitive view of causation, but propose a formal, modular framework that links causation-in-fact with legal relevance and accommodates multiple doctrinal approaches within a unified computational setting.

A relevant strand of work in AI takes the Halpern-Pearl model of actual causality as a starting point for defining notions of responsibility, harm and blame. Chockler & Halpern [10] change actual causation from an all-or-nothing concept (either c is a cause of e or it is not) into a gradual notion of responsibility. In case of multiple causes, this makes it possible to say to which degree each actual cause contributed to the effect. This idea could be relevant for defining degrees of liability, but that is not the same as determining whether an actual cause is legally relevant, as is our focus.

On the basis of their gradual notion of responsibility, Chockler & Halpern then define a notion of blame, which defines whether a person is to blame for some consequence in terms of what the person knew when performing the act or omission that actually caused the consequence. In essence, the person's degree of blame is the expected degree of responsibility given his or her subjective epistemic state (represented as a probability distribution over the possible situations before the action is performed). Chockler et al. [9] apply these ideas to various cases of determining legal causal responsibility, while Beckers et al. [6] include a notion of harm. Since the definitions of blame used in this body of work are epistemic, they do not yet capture moral or legal responsibility. Halpern & Kleiman-Weiner [14] further extend this work to notions of moral responsibility by adding the notion of intention, and Beckers [5] applies his variant of this approach to AI systems. It should be noted that the probability aspect in these approaches' definitions of blame is not the same as the legal notion of foreseeability, since the former is purely subjective while the latter is about what a reasonable person could foresee. More generally, this work does, contrary to us, not explicitly model a theory of legally relevant causation.

In terms of the five aspects of tortious liability which we distinguished in the introduction, we believe that this body of work primarily addresses aspects 3 (responsibility) and 4 (intention or negligence) of tortious liability while we primarily address aspect 2 (legally relevant cause-in-fact). In future work we want to investigate how this body of work can be connected to our model through the modular approach of [26], where each model address a different of the five aspects of tortious liability.

9 CONCLUSIONS

We have provided an argumentation-based model for legally relevant causality. As far as we know, this is the first computational model to cover this ground, also providing for a running implementation. The model uses the modular approach to argumentation of [26], to combine the determination of actual causation (causation-in-fact) with the further legal conditions required for causation to be legally relevant. Our instantiation of modular argumentation into distinct argumentation-based modules has the distinctive advantage of allowing defeasible causal rules, as well as defeasible legal rules on the relevance of causation, in ways that fit legal practice and the usual reasoning of lawyers.

In particular, causation in fact is based on a refined version of the argumentation-based approach provided in [23] while its legal relevance is based on a novel formalisation of the approach by [30]. We also show how our framework can be deployed for modelling different approaches and other aspects that, according to some doctrinal views, are included in the analysis of legal causality.

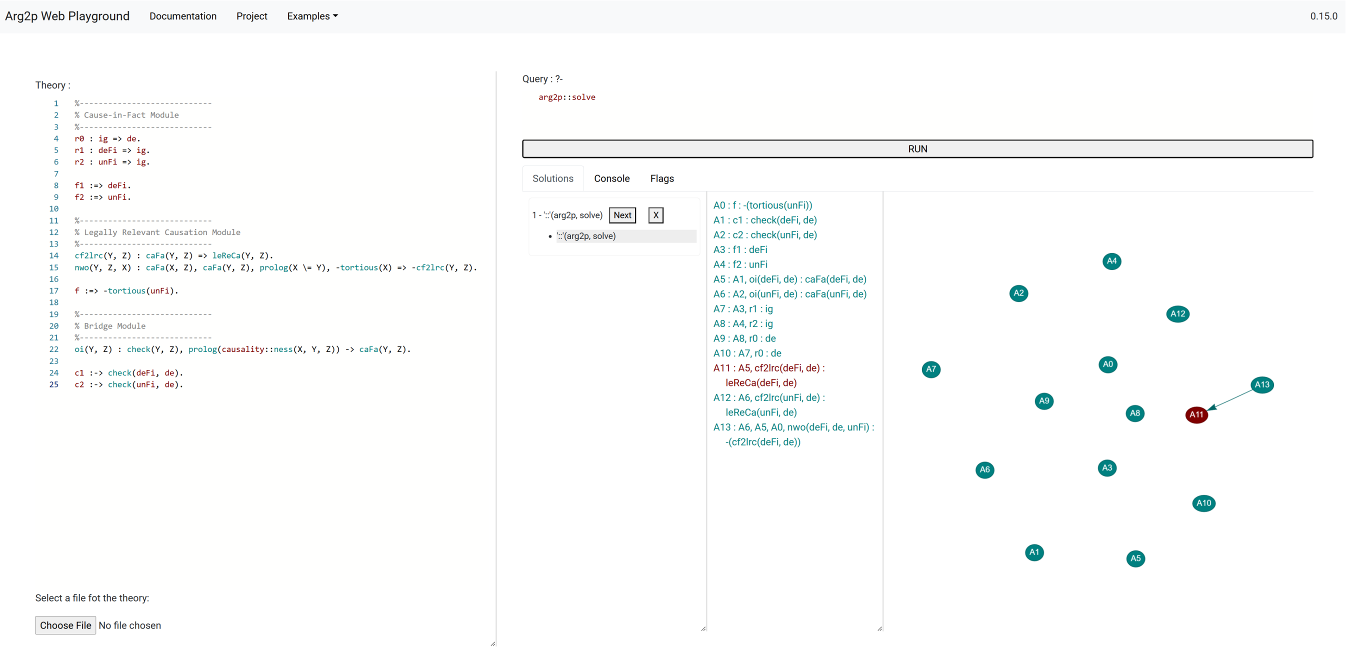
We have used our model of legally relevant causation to capture Wright's [30] analysis of such conditions, but it can be also applied to other doctrinal theories, as long as they allow for a clear distinction between natural causality and addition elements for it to be legally relevant (for an overview see [28]). We hope that our model can be useful both for theoretical studies on legal causation and for the analysis of concrete cases, supported by our (or others') implementations.

Acknowledgments

This work was supported by the European Research Council (ERC) under the Project "CompuLaw" (Grant Agreement No 833647) and by the Italian Ministry of University and Research (MUR) under the FIS project GenAI4Law (Grant No. FIS-2024-07323). We thank prof. Richard Wright for his very useful comments.

---

[4] https://tuprolog.github.io/arg2p-kt-web/



```
%---------------------------
% Cause-in-Fact Module
%---------------------------
r0 : ig => de.
r1 : deFi => ig.
r2 : unFi => ig.

f1 :=> deFi.
f2 :=> unFi.

%---------------------------
% Legally Relevant Causation Module
%---------------------------
cf2lrc(Y, Z) : caFa(Y, Z) => leReCa(Y, Z).
nwo(Y, Z, X) : caFa(X, Z), caFa(Y, Z), prolog(X \= Y), -tortious(X) => -cf2lrc(Y, Z).

f :=> -tortious(unFi).

%---------------------------
% Bridge Module
%---------------------------
oi(Y, Z) : check(Y, Z), prolog(causality::ness(X, Y, Z)) -> caFa(Y, Z).

c1 :-> check(deFi, de).
c2 :-> check(unFi, de).
```

```
arg2p::solve
```

```
1 - '::'(arg2p, solve)
  • '::'(arg2p, solve)
```

```
A0 : f : -(tortious(unFi))
A1 : c1 : check(deFi, de)
A2 : c2 : check(unFi, de)
A3 : f1 : deFi
A4 : f2 : unFi
A5 : A1, oi(deFi, de) : caFa(deFi, de)
A6 : A2, oi(unFi, de) : caFa(unFi, de)
A7 : A3, r1 : ig
A8 : A4, r2 : ig
A9 : A8, r0 : de
A10 : A7, r0 : de
A11 : A5, cf2lrc(deFi, de) :
   leReCa(deFi, de)
A12 : A6, cf2lrc(unFi, de) :
   leReCa(unFi, de)
A13 : A6, A5, A0, nwo(deFi, de, unFi) :
   -(cf2lrc(deFi, de))
```

Figure 1: Integrated *CIF* and *LRC* Theory Evaluation.

References

[1] H. Andreas, M. Armgardt, and M. Gunther. 2023. Counterfactuals for causal responsibility in legal contexts. *Artificial Intelligence and Law* 31 (2023), 115–132.

[2] M. Araszkiewicz. 2015. Incorporation of complex doctrinal theories in a model of statutory interpretation: an example of *adequate causal link*. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law, ICAIL 2015, San Diego, CA, USA, June 8-12, 2015*, Ted Sichelman and Katie Atkinson (Eds.). ACM, 161–165. doi:10.1145/2746090.2746114

[3] P. Baroni, M. Caminada, and M. Giacomin. 2011. An introduction to argumentation semantics. *The Knowledge Engineering Review* 26 (2011), 365–410.

[4] S. Beckers. 2021. The counterfactual NESS definition of causation. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)*. 6210–6217.

[5] S. Beckers. 2023. Moral responsibility for AI systems. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS '23)*. Curran Associates Inc., Red Hook, NY, USA.

[6] S. Beckers, H. Chockler, and J. Y. Halpern. 2022. A causal analysis of harm. In *Proceedings of the 36th International Conference on Neural Information Processing Systems (NeurIPS '22)*. Curran Associates Inc., Red Hook, NY, USA.

[7] F.J. Bex. 2011. *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory.* Springer, Dordrecht.

[8] R. Calegari, A. Omicini, G. Pisano, and G. Sartor. 2022. Arg2P: an argumentation framework for explainable intelligent systems. *Journal of Logic and ComputationComput.* 32, 2 (2022), 369–401.

[9] H. Chockler, N. Fenton, J. Keppens, and D.A. Lagnado. 2015. Causal analysis for attributing responsibility in legal cases. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*. ACM Press, New York, 33–42.

[10] H. Chockler and J. Y. Halpern. 2004. Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research* 22 (2004), 93 – 115.

[11] P.M. Dung. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence* 77 (1995), 321–357.

[12] J. C.P. Goldberg and B. C. Zipurksy. 2016. The strict liability in fault and the fault in strict liability. *Fordham Law Review* 85 (2016), 744–86.

[13] J.Y. Halpern. 2016. *Actual Causality.* MIT Press.

[14] J. Halpern and M. Kleiman-Weiner. 2018. Towards formal definitions of blameworthiness, intention, and moral responsibility. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*. https://doi.org/10.1609/aaai.v32i1.11557.

[15] J.Y. Halpern and J. Pearl. 2005. Causes and explanations: A structural-model approach. Part I: Causes. *Brit. J. Phil. Sci.* 56 (2005), 843–887.

[16] H.L.A. Hart and T. Honoré. 1985. *Causation in the Law.* OUP.

[17] W. Landes and R. A. Posner. 2003. *The Economic Structure of Intellectual Property Law.* Belknap.

[18] J. Lehmann and A. Gangemi. 2007. An ontology of physical causation as a basis for assessing causation in fact and attributing legal responsibility. *Artificial Intelligence and Law* 15 (2007), 301–321.

[19] R. Liepina, T. de Lima, E. Lorini, G. Pisano, and G. Sartor. 2025. A causal model checker for legal cases. In *Proceedings of the 20th International Conference on Artificial Intelligence and Law.* ACM Press, 744–786.

[20] S. Modgil and H. Prakken. 2013. A general account of argumentation with preferences. *Artificial Intelligence* 195 (2013), 361–397.

[21] S. Modgil and H. Prakken. 2014. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation* 5 (2014), 31–62.

[22] S. Modgil and H. Prakken. 2018. Abstract rule-based argumentation. In *Handbook of Formal Argumentation*, P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre (Eds.). Vol. 1. College Publications, London, 286–361.

[23] G. Pisano, H. Prakken, G. Sartor, and R. Liepina. 2025. Modelling cause-in-fact in legal cases through defeasible argumentation. In *Proceedings of the 20th International Conference on Artificial Intelligence and Law.* ACM Press, 278–287.

[24] H. Prakken. 1999. Dialectical proof theory for defeasible argumentation with defeasible priorities (preliminary report). In *Formal Models of Agents (Springer Lecture Notes in AI, 1760)*, J.-J.Ch. Meyer and P.-Y. Schobbens (Eds.). Springer Verlag, Berlin, 202–215.

[25] H. Prakken. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1 (2010), 93–124.

[26] H. Prakken and G. Sartor. 2023. A formal famework for combining legal reasoning methods. In *Proceedings of the 19th International Conference on Artificial Intelligence and Law.* ACM Press, 227–236.

[27] R.W. Wright. 2011. The NESS account of natural causation: a response to criticisms. In *Perspectives on Causation*, R. Goldberg (Ed.). Hart Publishing, Chapter 14.

[28] R.W. Wright and I. Puppe. 2016. Causation: linguistic, philosophical, legal and economic. *Chi.-Kent L. Rev.* 91 (2016), 461.

[29] R. W. Wright. 2003. The Grounds and Extent of Legal Responsibility. *San Diego Law Review* 40 (2003), 1425–1532.

[30] R. W. Wright. 2025-2006. Causation and Legal Responsibility. In *CTLA Forum 2025-2026. CTLA Forum, Winter 2025 at 18-22, 40-45, Spring 2025 at 40-45, 50-54.* https://www.academia.edu/127567497/Causation_and_Legal_Responsibility

[31] S. Yablo. 2010. Advertisement for a sketch of an outline of a proto theory of causation. In *Things: Papers on Objects, Events, and Properties.* OUP, 98–116.