Modelling Cause-in-Fact in Legal Cases through Defeasible Argumentation

Giuseppe Pisano
University of Bologna, Law Faculty-CIRSFID
Italy
g.pisano@unibo.it

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

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

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

Abstract

We propose to model cause-in-fact in legal cases through fresh argumentation-theoretic notions of explanation and support, meant to capture the set of arguments that contribute to making a conclusion justified. This novel argumentation-based approach to causality in law goes beyond the traditional idea of a cause as a necessary antecedent condition (the *conditio-sine-qua-non* idea), to handle concurrent causal processes leading to overdetermination and preemption. It also provides sound analyses of cases involving omission and ennoblement. Finally, by relying on defeasible argumentation it can capture causal inferences based on defeasible generalisations, which are very often used in judicial reasoning. Through the analysis of causal puzzles in legal cases, we illustrate the framework's effectiveness in handling complex causal reasoning, and demonstrate its potential to support legal reasoners with structured and intuitive analysis.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**;

Keywords

Causation in law, Cause-in-fact, Argumentation

ACM Reference Format:
Giuseppe Pisano, Henry Prakken, Giovanni Sartor, and Rūta Liepiņa. 2025. Modelling Cause-in-Fact in Legal Cases through Defeasible Argumentation. In *Twentieth International Conference on Artificial Intelligence and Law (ICAIL 2025), June 16–20, 2025, Chicago, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

ICAIL 2025, June 16–20, 2025, Chicago, USA

ACM ISBN 979-8-4007-0197-9/23/06
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Causal reasoning is important in the law in various ways. It can address different problems concerning one or more antecedent facts P and a subsequent fact Q. First, where both P and Q are true, we may want to know whether the causal rule 'P causes Q' should be part of our causal theory T (e.g.: Is it true that P causes disease Q?). This is sometimes called *causal discovery*. Then, were Q is true and a causal theory T is given, we may want to know which of the $P's$ that could cause Q according to T is true. This is often called *abduction* or *inference to the best explanation* (e.g.: Can we infer from evidence Q that the unlawful action P was accomplished by the accused?). Next, where P is true and a causal theory T is given, we may want to know what else will be true because of P. This is sometimes called *causal prediction* (e.g.: Is it true that the enactment of legislative act P will reduce unemployment, Q?). Finally, where both P and Q are true and a causal theory T is given, we may want to know whether P is a cause of Q according to T. This is sometimes called the problem of *actual causation* (e.g.: Given facts and causal rules, can we conclude that action P by the accused caused the harmful event Q?). This paper is about actual causation, which in the law is often referred to as cause-in-fact.

Actual causation is highly relevant to law, in particular where liability for a harmful event has to be attributed, though further elements may be needed for liability to be established: (1) the harm must have been caused in a way that is relevant to the law (2) the liable agent must be in an appropriate relation to the harm-causing event (being the originator of the event or being responsible for the thing or person originating it), (3) there must be an appropriate mental state in the agent, if required by the law (intention or negligence, except for cases of strict liability), (4) there must not exist circumstances that exclude liability (such as self defence).

Since legal liability is such a complex issue, it is necessary to precisely circumscribe the scope of our contribution. We only consider actual causality, without entering into the further conditions of liability just listed from (1) to (4). We also refrain from investigating any issues pertaining to the proof of causality. Finally, we only address the core causality in law, namely the idea of actual causality, to which we refer as cause-in-fact.

Following Richard W. Wright [32, 33], we indeed argue that clarity and interdisciplinary insights in the analysis of legal causality

(for a review, see [22]) could be obtained by carefully distinguishing and separately theorising a basic and general cross-domain notion of cause-in-fact and any further condition for the legal relevance of a cause-in-fact. Thus, to address the legal significance of casual connections, two steps are needed.

First, a notion of cause-in-fact is required, to precisely identify causal relations. This notion must be subtle enough to correctly and precisely address all instances of actual causality. In particular, it must successfully address the following controversial aspects pertaining to causation in law (but also to other domains): i) *overdetermination*, when 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) *omission*, where the effect is caused by the non-occurrence of an event; and iv) *ennoblement*, where both the occurrence and non-occurrence of an event would lead to an effect, through different causal paths (see [34]) .

Secondly, conditions have to be specified which are required for a cause-in fact to generate legal liability (e.g., a certain kind of *ex ante* predictability), and, alternatively or additionally, exceptions have to be specified under which a cause-in-fact fails to generate such responsibility (e.g., the intervention of an extraordinary event leading to unexpected effects). These conditions and exception may differ in different legal systems.

We shall only consider the first step, i.e., the general idea of a cause-in-fact. Thus we do not commit to whether the conditions required for a cause-in fact are to be legally relevant and the exceptions excluding such relevance are to be included (in addition to cause-in-fact) within the concept of a cause in law (e.g., in the idea of a legally proximate or adequate cause) or are to be viewed as separate requirements for liability (as in the third restatement of the US law on torts, see [33, 464]). However, we will mention that both perspectives could be captured within an argumentation-based model.

Our model of cause-in-fact will be formulated within formal argumentation. It is 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 the E to be produced ($S/\{A\}$ will not be sufficient to produce the effect). This approach, introduced in the seminal contribution by Hart and Honoré [16], has been further developed by [33] and has recently been endorsed in AI by, among others, Halpern [14] and Beckers [3]. A strong point of NESS is that it 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.

Our choice for an argumentation approach is motivated by structural similarities between the NESS model of causality and argumentation. First, as NESS admits the coexistence of multiple different sufficient sets of conditions, so in argumentation a conclusion may be supported by multiple parallel arguments. Secondly, as NESS requires that each element in a sufficient set is necessary, so in argumentation, arguments must be minimal, i.e., they cannot contain redundant premises. Third, for the set of conditions to be sufficient, all exceptions pre-empting the generation of the effect have to be overcome. Similarly in argumentation, for an argument to be justified, it must be able to survive all attacks by being defended by other arguments. Finally, each argument can have multiple ways of being defended against attacks, and each of these ways can be, on its own, sufficient to defend the argument against an attack, while each element in such a defence is necessary for the defence to have its effect.

We believe that an argumentation-based approach has some advantages, in the legal domain, in comparison to other leading approaches to actual causality, such as the approach with Structural Causal Models of [15], which however has been a key inspiration for our work. In particular, our approach, being based upon defeasible argumentation, has the advantage of directly addressing defeasible causality, which plays a key role in legal reasoning on matters of fact (where common-sense and other defeasible generalisation play a key role). Moreover, the argumentation approach directly models arguments aimed at supporting causal claims or at rejecting them. Thus it can be more easily mapped into real instances of legal reasoning and understood by a legal audience. More specifically, we will use a combination of the theory of abstract argumentation frameworks [11] with the *ASPIC*$^+$ framework for argumentation; this combination provides the formal tools we need.

To demonstrate the effectiveness of our argumentation-based approach –as a refinement and formalisation of the NESS idea– and showcase its ability to address causal puzzles (involving overdetermination, pre-emption, omission and ennoblement) we provide various examples from judicial cases and from the literature.

The paper is structured as follows. First, in Section 2 we introduce the theory of abstract argumentation frameworks and the *ASPIC*$^+$ approach to argumentation. In Section 3, we propose a formal notion of relevance for abstract argumentation frameworks. In Section 4, we embed this notion in *ASPIC*$^+$ and use this embedding to formalise an argumentation-based notion of cause-in-fact. In section 5, we apply this notion to several legal cases, showing how our formalisation successfully deals with the above-mentioned issues (overdetermination, confounding, preemption and omission). In Section 6, we formalise some of these examples in an alternatIve way inspired by the Event Calculus, to capture temporal aspects in a more general way. In Section 7, we discuss related work, after which we conclude in Section 8.

2 FORMAL PRELIMINARIES

In this section we present our formal preliminaries, being the theory of abstract argumentation frameworks [11] and the *ASPIC*$^+$ framework for structured approaches to argumentation [21].

2.1 Abstract Argumentation Frameworks

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 defeat.[1] 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 $\mathcal{D} = \{(B, A), (B, C), (C, B)\}$. Let $S \subseteq A$. Then S is conflict-free if no member of S defeats a member of S and S

[1] Dung used the term 'attack' but since we want to instantiate it with the *ASPIC*$^+$ defeat relation, we rename it to 'defeat'.

(directly) *defends* $A \in \mathcal{A}$ if for all $B \in \mathcal{A}$: if B defeats A, then some $C \in S$ defeats B.

A set $S \subseteq \mathcal{A}$ is *admissible* if it is conflict-free and defends all its members. A set $S \subseteq \mathcal{A}$ is *strongly admissible* [10] if every $A \in S$ is defended by some $S' \subseteq S \setminus \{A\}$ which is strongly admissible. The intuitive difference between the latter two notions is that for admissibility an argument can defend itself while for strong admissibility an argument has to be defended by another argument. For example, in case of $A \leftrightarrow B$, all of $\emptyset$, $\{A\}$ and $\{B\}$ are admissible while only $\emptyset$ is strongly admissible.

The semantics of *AFs* [2, 11] identifies sets of arguments (called *extensions*) which are internally conflict-free (no member attacks a member) and defend themselves against all attackers. In this paper we use labelling-style semantics. A *labelling* of a set $\mathcal{A}$ 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* (for 'undecided') iff it is neither *in* nor *out*.

In this paper we focus on grounded semantics, leaving generalisation of our approach to other semantics for future research. 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. We say that an argument A is *justified*, respectively, *defensible*, *overruled* if A is *in*, respectively, *und*, *out* in the grounded labelling.

To briefly illustrate these notions, in case of $A \leftrightarrow B$ all arguments are undecided so defensible, so the grounded extension is the empty set, while in case of $A \leftarrow B \leftarrow C$ arguments A and C are *in* while B is *out*, so A and C are justified while B is overruled, and the grounded extension is $\{A, C\}$.

It is known that an argument A is in the grounded extension of AF if and only if the proponent in the so-called *grounded argument game* has a winning strategy for A [23]. Briefly, in the grounded argument game a proponent and an opponent of an argument A exchange arguments, taking turns after each move and picking all their arguments 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 that player can make the other player run out of moves in whatever way the other player plays. A winning strategy for player p can be visualised as a tree with A as root and in which all branches are terminated games won by p, where p's moves have as children all possible replies of the other player, $\overline{p}$ and where $\overline{p}$'s moves have exactly one child.

Consider the example AF in Figure 1, which shows the grounded labelling. The grounded extension is $\{A, D, F, G\}$. To prove that A is in the grounded extension, the proponent has just one winning strategy, namely, $A \leftarrow B \leftarrow D$, since if the proponent responds to B with C, then the opponent can win by moving F.

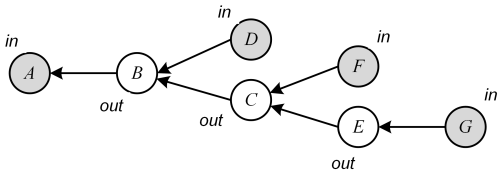

Figure 1: (Ir)relevant arguments (1)

We finally recall how Borg & Bex [9] recursively define the set of defenders (whether directly or indirectly) of an argument.

Definition 2.1. [**Defenders [9]**] Let $AF = (\mathcal{A}, \mathcal{D})$ and $A, B \in \mathcal{A}$. Then B is a *direct defender* of A iff $(B, C) \in \mathcal{D}$ for some $C \in \mathcal{A}$ such that $(C, A) \in \mathcal{D}$. And B is an *indirect defender* of A iff for some $C \in \mathcal{A}$ it holds that C is a (direct or indirect) defender of A and B is a (direct or indirect) defender of C.

In Figure 1 arguments C and D directly defend A while G indirectly defends A.

2.2 The *ASPIC*$^+$ Framework

The *ASPIC*$^+$ framework [20, 21, 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 $\mathcal{L}$. Over the years, several variants of the framework have been developed. The version we use is a special case of the framework of [25], which sufficies for our purposes. See section 2.2 of [26] for a discussion of other variants of *ASPIC*$^+$.

In this paper we for simplicity assume that $\mathcal{L}$ contains ordinary negation $\neg$ but all new definitions proposed in this paper can be easily adapted to versions of *ASPIC*$^+$ with asymmetric negation, such as negation as failure. Arguments are constructed from a knowledge base (a subset of $\mathcal{L}$) by chaining inferences over $\mathcal{L}$ into acyclic graphs.

Definition 2.2. [**Argumentation System**] an *argumentation system* (AS) is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language with a negation symbol $\neg$;
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ 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 wff 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 a well-formed 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 2.3. [**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 2.4. [**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.

We introduce a running example to illustrate our framework. We model it through a combination of exogenous facts (facts that are not conclusions of rules) and causal rules according to which certain facts cause other facts. Note that our causal rules are both contextual and defeasible, as is usual for most common-sense causal connections put forward in legal cases.

Example 2.5 (Suzy throws a stone). Suzy decides to throw a stone at a window and implements her decision. The stone hits the window, which shatters into pieces. Her friend Billy is with her. He could stop her (by blocking her arm) but does not do it.

Let us use the following atoms: *SuDe* = Suzy decides to throw the stone; *SuTh* = Suzy throws the stone; *SuHi* = Suzy hits the window; *WiSh* = the window is shattered; *BiSt* = Billy stops Suzy. We can model our example through an argumentation theory AT_1 with

$$\begin{array}{lll} \mathcal{R}_1 = & \{ & r_1 : SuDe \Rightarrow SuTh; \quad r_2 : SuTh \Rightarrow SuHi; \\ & & r_3 : SuHi \Rightarrow WiSh; \quad r_4 : BiSt \Rightarrow \neg r_1\} \\ \mathcal{K}_1 = & \{ & SuDe\} \end{array}$$

Our knowledge base says the following: Suzy decides to throw the stone (*SuDe*), by r_1 if she decides to throw she does it (*SuTh*), by r_2 if she throws she hits the window (*SuHi*), by r_3 if she hits the window the window is shattered (*WiSh*), by r_4 if Billy stops Suzy (*BiSt*) then it is not the case that by if Suzy decides to throw she does it (r_1 does not apply).

Definition 2.6. [**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) $A_1, \ldots, A_n \rightsquigarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that $\psi \notin \mathtt{Conc}(\{A_1, \ldots, A_n\})$ and $\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \rightsquigarrow \psi \in \mathcal{R}$ with:
$\mathtt{Prem}(A) = \mathtt{Prem}(A_1) \cup \ldots \cup \mathtt{Prem}(A_n)$;
$\mathtt{Conc}(A) = \psi$;
$\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n) \cup \{A\}$;
$\mathtt{Rules}(A) = \mathtt{Rules}(A_1) \cup \ldots \cup \mathtt{Rules}(A_n) \cup \{\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \rightsquigarrow \psi\}$;
$\mathtt{DefRules}(A) = \mathtt{Rules}(A) \cap \mathcal{R}_d$;
$\mathtt{TopRule}(A) = \mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \rightsquigarrow \psi$.

$\mathtt{Prem}_n(A) = \mathtt{Prem}(A) \cap \mathcal{K}_n$ 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 Func in this definition is also defined on sets of arguments $S = \{A_1, \ldots, A_n\}$ 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.

Example 2.7 (Suzy throws a stone - Arguments). Given the argumentation theory AT_1 of Example 2.5 we can build the following arguments (also displayed on the left side of Figure 2).

$$\begin{array}{ll} A_1 = SuDe; & A_2 = A_1 \Rightarrow_{r_1} SuTh; \\ A_3 = A_2 \Rightarrow_{r_2} SuHi; & A_4 = A_3 \Rightarrow_{r_3} WiSh \end{array}$$

Definition 2.8. [**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 $B'_1, \ldots, B'_n \Rightarrow \varphi$.
- A *undermines* B (on φ) iff $\mathtt{Conc}(A) = -\varphi$ for some $\varphi \in \mathtt{Prem}(B) \cap \mathcal{K}_p$.

Example 2.9 (Suzy throws a stone and Billy blocks her). Consider an argumentation theory with $\mathcal{R}_2 = \mathcal{R}_1$ and $\mathcal{K}_2 = \mathcal{K}_1 \cup \{BiSt\}$. Then, in addition to the arguments in Example 2.7, we have arguments $A_5 = BiSt$ and $A_6 = A_5 \Rightarrow_{r_4} \neg r_1$ (also displayed in Figure 2). Argument A_6 undercuts A_2, A_3 and A_4—i.e., since Billy stops Suzy from throwing, she does not hit and, consequently, the window does not shatter.

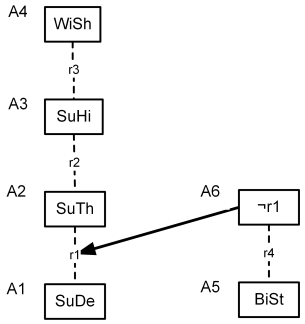

Figure 2: Undercutting attack

The notion of *defeat* is now defined as follows. Undercutting attacks succeed as *defeats* independently of preferences over arguments, since they express exceptions to defeasible rules. Rebutting and undermining attacks succeed only if the attacked argument is not stronger than the attacking argument, where $A \prec B$ is defined as usual as $A \preceq B$ and $B \not\preceq A$ and $A \approx B$ as $A \preceq B$ and $B \preceq A$. Below we assume the so-called basic argument ordering, according to which if A is strict and firm while B is either defeasible or plausible, then $A \prec B$ does not hold, so a rebutting attack of A on B always succeeds as defeat. For present purposes this is all we need.

Definition 2.10. [**Defeat**] Argument A *defeats* argument B iff either A undercuts B; or A rebuts B on B' and $A \not\prec B'$.

Definition 2.11. [**Structured Argumentation Frameworks**] A *structured argumentation framework (SAF)* defined by an argumentation theory AT is a triple $(\mathcal{A}, \mathcal{C}, \preceq)$ where $\mathcal{A}$ is the set of all arguments on the basis of AT, $\preceq$ is the basic ordering on $\mathcal{A}$, i.e., $A \prec B$ iff A is defeasible or plausible and B is strict and firm, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

Abstract argumentation frameworks are then generated from *SAFs* as follows:

Definition 2.12. [**Argumentation frameworks**] An *abstract argumentation framework (AF) corresponding to a SAF* = $(\mathcal{A}, C, \preceq)$ is a pair $(\mathcal{A}, \mathcal{D})$ such that $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by *SAF*.

Example 2.13 (Suzy throws a stone and Billy stops her - Argumentation framework). Given the *SAF* defined by AT_2 from Example 2.9 and the basic argument ordering, the argumentation framework AF_{AT2}[2] is shown in Figure 3.

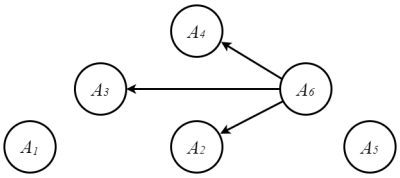

Figure 3: Argumentation framework from Example 2.9.

Finally, we say that $\varphi \in \mathcal{L}$ is *justified* on the basis of a *SAF* if φ is the conclusion of a justified argument on the basis of the *AF* corresponding to the *SAF* under semantics T, and *defensible* if φ is not justified and is the conclusion of a defensible argument.

3 RELEVANCE IN ARGUMENTATION

To exploit the structural similarity between argumentation-theoretic notions of relevance and the notion of actual causation, we next discuss what is a suitable way to formalise argumentation-theoretic notions of relevance. Borg & Bex [9] give definitions that, when applied to grounded semantics, essentially define the set of all arguments relevant to A in the grounded extension $\mathcal{E}$ (which we will call the *support set* of A relative to $\mathcal{E}$) as the set of all arguments in $\mathcal{E}$ that are (direct or indirect) defenders of A. However, this definition is too broad for our purposes, as the following examples show. The following example (taken from [24]) shows that defeating a defender of A does not necessarily change the justification status of A.

Example 3.1. In the *AF* in Figure 1, the grounded extension is $\{A, D, F, G\}$. Note that C and G are defenders of A but defeating either of them does not lower the status of A; this only happens if either A or D is defeated. So intuitively C and G should not be regarded as relevant to A. Yet they both are in the sense of relevance defined by [9].

Fan & Toni [12] define notions of an explanation in terms of so-called *related admissible sets*. Given an $AF = (\mathcal{A}, \mathcal{D})$, a set $S \subseteq \mathcal{A}$ is *related admissible* if there exists an $A \in S$ such that S 'defends' A and S is admissible. Here S 'defends' A iff for all $B \in S$ it holds that either $B = A$ or B is a defender of A as defined in Definition 2.1 above. Then any related admissible set containing A is an *explanation* of A, while any subset-minimal related admissible set containing A is a *compact explanation* of A. These notions may be suitable for preferred semantics, since it is known that every admissible set is contained in a preferred extension. However, for grounded semantics they are too broad, as the following example shows.

[2]This notation stands for the *AF* corresponding to the *SAF* defined by *AT*2 and the basic argument ordering.

Example 3.2. Consider the *AF* in Figure 4. The grounded extension is $\{A, C, D\}$. We have that $\{A, C\}$ is a compact explanation of A and a subset of the grounded extension. Yet intuitively, C should not be in the support set of A since A is in the grounded extension because of D: if D is deleted from the *AF* then A is not in the grounded extension any more.

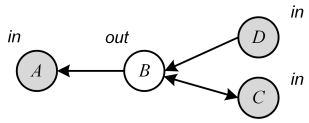

Figure 4: (Ir)relevant arguments (2)

For finite *AF* a satisfactory solution to this problem is to define the support set of $A \in \mathcal{E}$ relative to the grounded extension $\mathcal{E}$ as the set of all arguments that are in any subset-minimal *strongly* admissible subset of $\mathcal{E}$ containing A, since for finite *AF* the grounded extension is equal to the unique maximal strongly admissible subset of $\mathcal{A}$ [10]. Note that in Example 3.2, $\{A, D\}$ is the only subset-minimal strongly admissible subset of $\mathcal{E}$ containing A. However, for infinite *AF* this solution does not work, since for the infinite case the grounded extension is not always equal to the subset-maximal strongly admissible set. A fully general solution is to define support sets of an argument A for grounded semantics in terms of minimal winning strategies for A in the grounded game (minimal in that the set of proponent arguments in the winning strategy is not a strict superset of the set of proponent arguments in any other winning strategy for A). Henceforth we will call any such set an *explanation* of A. Formally:

Definition 3.3. [**Explanations, support sets, and relevance**] An *explanation* for an argument A being in the grounded extension $\mathcal{E}$ of an argumentation framework $AF = (\mathcal{A}, \mathcal{D})$ is the set of all proponent arguments in any minimal winning strategy for A in the grounded argument game.

The *support set* $Supp(A, \mathcal{E})$ of argument A relative to extension $\mathcal{E}$ is the union of all explanations of A being in $\mathcal{E}$. For any formula ψ which is the conclusion of an argument in $\mathcal{E}$, the *support set* $Supp(\psi, \mathcal{E})$ of ψ relative to $\mathcal{E}$ is the union of all support sets $Supp(A, \mathcal{E})$ of any argument $A \in \mathcal{E}$ with conclusion ψ.

An argument B is *positively relevant* to argument A with regard to $\mathcal{E}$ iff $B \in Supp(A, \mathcal{E})$, and B is positively relevant to formula ψ with regard to $\mathcal{E}$ iff $B \in Supp(\psi, \mathcal{E})$.

That winning strategies must be minimal ensures that any successful attack on an argument positively relevant to A makes the proponent lose at least one way to show that A is justified.

4 AN ARGUMENTATION-THEORETIC DEFINITION OF CAUSE-IN-FACT

In this section, we combine our concept of relevance in abstract argumentation frameworks with *ASPIC*$^+$ in order to give an argumentation-theoretic definition of cause-in-fact. Our definition assumes that we represent causal relations as defeasible rules and facts as ordinary premises. Suppose ψ is justified and let S be the support set of ψ (so the union of all support sets of all justified arguments for ψ). We

then want to define the causes of ψ. We use the following example to illustrate our definitions.

Example 4.1 (Cause). Consider an argumentation theory where rules $\mathcal{R}$ and facts $\mathcal{K} = \mathcal{K}_p$ are:

$\mathcal{R}=$ $\{r_1 : p \Rightarrow q, r_2 : r \Rightarrow \neg r_1, r_3 :\Rightarrow \neg r_2, r_4 : t \Rightarrow \neg r_3\}$
$\mathcal{K}=$ $\{p, r, s, \neg t\}$

Then q is justified on the basis of AF_{AT}. The arguments in AF 'about' q are $A = p \Rightarrow q$, $B = r \Rightarrow \neg r_1$ and $C = s \Rightarrow \neg r_2$. The support set of q is the support set of $A = \{A, C\}$.

Now the causes of ψ should surely include all formulas in any justified argument for ψ. Note that these are exactly all conclusions of arguments in the support set of ψ, namely all conclusions of arguments that are positively relevant to ψ (see Definition 3.3). So in Example 4.1 we want that p is a cause of q. We also want that any formula in any (direct or indirect) defender of any argument for ψ is a cause. So in Example 4.1 we want that s is a cause of q. However, our notion of a cause goes beyond what is positively relevant. We also need to capture what is 'negatively' relevant, i.e., those propositions such that their complement would contribute to a counterargument that would prevent an argument in the support set from being justified. In other terms, if proposition φ would contribute to (would be included in) a counterargument that would defeat an argument in the support set for ψ, then $-\varphi$ is a cause of ψ. So in Example 4.1 we want that $\neg t$ is a cause of q since adding t to $\mathcal{K}$ would create a new undefeated undercutter of argument C. Note that the negations of some of these potential undercutters may be in the current AT, either as an ordinary premises or as a conclusion of a defeasible-rule application (in our example $\neg t$ is in $\mathcal{K}_p$). For this reason the additions to $\mathcal{K}$ should be necessary premises, so that they can be used to strictly defeat both ordinary premises and defeasible conclusions. So in Example 4.1, since $\neg t$ is in $\mathcal{K}_p$, we want to add t to $\mathcal{K}_n$.

Accordingly, we define an axiom-expansion of an AT as follows.

Definition 4.2. [**Axiom-expansions**] The *axiom-expansion* $\mathcal{K}+L$ of a knowledge base $\mathcal{K} = \mathcal{K}_p \cup \mathcal{K}_n$ is defined as $\mathcal{K} + L = \mathcal{K}_p \cup (\mathcal{K}_n \cup L)$, where $L \subseteq \mathcal{L}$ is a set of axioms. The *axiom expansion of an argumentation theory* $AT = (AS, \mathcal{K})$ with L is defined as $AT + L = (AS, \mathcal{K} + L)$.

The axiom-expansion of $\mathcal{K}$ with L provides indefeasible arguments for all atoms in L (since the elements of L are considered necessary premises), defeating (and thus making irrelevant) all arguments that contain $-L$ as a premise or conclusion. Thus, it corresponds, within our argumentation framework, to an *intervention* as defined by [15, 487], which consists in modifying a causal model by setting the values of certain variables and removing all their dependencies from other variables.

We can now give our formal definition of a cause. As for notation, $\mathcal{G}(AF)$ denotes the grounded extension of AF.

Definition 4.3. [**Cause**] Literal φ is a *cause* of literal ψ relative to AF_{AT} iff there exist a subset-minimal set of literals L and an explanation E for argument $A \in \mathcal{G}(AF_{AT})$ with $Conc(A) = \psi$ such that

- $-\varphi \in L$,
- $E \not\subseteq \mathcal{G}(AF_{AT+L})$

In other words, for φ to be a cause of ψ it is required that by adding literals L to the knowledge base we obtain a non-overruled defeater (containing $-\varphi$) of an explanation of ψ. Thus, the test for whether φ is a cause of ψ amounts to adding its contradictory $-\varphi$ as a necessary premise (which thus overrides any incompatible ordinary fact or defeasible conclusion) and then checking whether we have lost an explanation for ψ. According to this definition we can indeed distinguish three ways in which –given an explanation E for an argument A having conclusion ψ– a literalφ can be a cause of ψ:

(1) φ is the conclusion of a subargument of A. In this case, $-\varphi \in L$ leads to A being rebutted or undermined by $-\varphi$ in AF_{AT+L}. This attack succeeds as an undefeated defeat, since $-\varphi \in \mathcal{K}_n$, so $A \notin \mathcal{G}(AF_{AT+L})$, and E is no explanation of A relative to AF_{AT+L};
(2) φ is included in a justified argument $C \in E$ defending an argument A for ψ. Then $-\varphi \in L$ leads to C being excluded from $\mathcal{G}(AF_{AT+L})$ for the same reason, so E is no explanation of A relative to AF_{AT+L};
(3) $-\varphi \in L$ enables or defends a new argument $B \in AF_{AT+L}$ that defeats an argument $C \in E$ such that $C \notin \mathcal{G}(AF_{AT+L})$. Then, too, E is no explanation of A relative to AF_{AT+L}.

In Example 4.1 we have that p is a cause under (1); s is a cause under (2) and $\neg t$ is a cause under (3). Note that $\neg t$ would also be a cause under (3) if $\neg t$ were not in $\mathcal{K}$. Note also that according to Definition 4.3 also certain undercutting conclusions (i.e., negations of rule names) would count as causes. For instance, in example 4.1 $\neg r_2$ would count as a cause. Such conclusions play a technical role in our model of argument (since they lead to undercutting other arguments), but it may be less natural to include them among the causes-in-fact. They can be filtered out if desired, for instance, by requiring that causes appear in the antecedents of rules.

Our notion of a cause can be directly connected to the basic concept of NESS, namely, the idea that a cause is a necessary element of a set of elements that is sufficient for the effect to happen. Remember that the support set for a proposition is the union of all explanations for that proposition, and that the explanations of an argument A are minimal sets of arguments that ensure that A is justified. Thus any cause ϕ of ψ is necessary for one particular explanation (justification) E of ψ to hold:

- ϕ could be a conclusion of an argument in an explanation E for ψ, and would therefore be necessary for E to exist (cases (1) and (2) above), or
- $-\phi$ could lead to a successful new challenge against an argument in E, so that E no longer explains (justifies) ψ (case (3) above). Thus ϕ is necessary for E to explain ϕ because it is needed to repel that challenge.

It may seem that this concept of causation leads to too many antecedent facts being qualified as causes of the effect at stake. However, the problem context will always yield just a small set of legally relevant potential causes (typically, the actions/omissions of the person whose responsibility is being considered, or the outcome of such an action).

5 CAUSE-IN-FACT IN LEGAL CASES

In this section, we apply our framework to analyse cause-in-fact in six legal cases. Each scenario consists of a description of the case, the variables, and the causal model.

Asbestos and Lung Cancer (Overdetermination): Fairchild v Glenhaven (2002). [3]

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).

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

Each Ex_i (with $i = 1, 2, 3$), i.e., the exposure while working in company i, is a cause of Di, since each Ex_i is included in an explanation for Di. Just take any argument for Di which includes subargument Ex_i, for instance, $Ex_1, Ex_2 \Rightarrow Ca$ is such an argument.

Double Shooting, Attempted Crime (Preemption): People v. Dlugash (1997). [4] In 1973, Mr. Geller was found shot to death in his Brooklyn apartment. Bush, had shot first and fatally wounded the victim whereas Dlugash had fired his shots after Geller was dead. 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 but failed to do so. We provide two models of this example. The first simpler one, does not include a representation of time, which is used in the second, in Section 6 below. Here are non-temporalised atoms: Bush/Dlugash shoots ($BuSh/DlSh$); Bush/Dlugash kills Geller ($BuKi/DlKi$); Geller dies ($GeDi$). This model deals with preemption, by explicitly stating that if Bush has killed Geller, then Dlugash shot cannot have this effect.

$$\begin{array}{lll} \mathcal{R} = & \{ & r_0 : BuSh \Rightarrow BuKi, \quad r_1 : DlSh \Rightarrow DlKi, \\ & & r_2 : BuKi \Rightarrow GeDi, \quad r_3 : DlKi \Rightarrow GeDi, \\ & & r_4 : BuKi \Rightarrow \neg r_1\} \\ \mathcal{K} = & \{ & BuSh, DlSh\} \end{array}$$

It is easy to see that Bush shooting ($BuSh$) is a cause of Geller's dying ($GeDi$) since it is included in a justified argument leading to that effect. Dlugash's shooting ($DlSh$) is not since the argument for $GeDi$ including $DlSh$ is undercut according to rule r_4, i.e., by the justified argument $[BuSh \Rightarrow BuKi] \Rightarrow \neg r_1$ [5]

Childhood Leukaemia (Omission): Cassazione penale (2023). [6]

This Italian case 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 was not treated and died soon after. The parents were held responsible under civil law for causing the death of the child. Here are the atoms: child has leukaemia ($ChLe$); child dies ($ChDi$); child receives chemotherapy ($Chem$); parents consent ($PaCo$).

$$\begin{array}{lll} \mathcal{R} = \{ & r_0 : ChLe \Rightarrow ChDi, & r_1 : PaCo \Rightarrow Chem, \\ & r_2 : Chem \Rightarrow \neg r_0\}, & \\ \mathcal{K} = \{ & ChLe, \neg PaCo\} & \end{array}$$

Leukaemia ($ChLe$) is a cause of $ChDi$ but $\neg PaCo$ also is a cause. In fact, we can build a refutation of $ChDi$ by adding the complement of $\neg PaCo$, i.e., $PaCo$ to the knowledge base. This intervention enables the following justified argument : $[PaCo \Rightarrow Chem] \Rightarrow \neg r_0$. This counterfactual argument says that if the parents had consented then causal rule r_0 (according to which leukaemia will lead to the child's death) would not not have applied.

Car Accident (Preemptive Negative Causation): Saunders v Adams (1928). [7] A car ran into a motorist. The car driver did not press the brake pedal, but the brakes were defective so even if the driver had tried to brake, the accident would have happened anyway. According to the causal model endorsed by judges [32], the driver was considered to have caused the crash. Atoms are: accident happens ($AcHa$); brakes fail ($BrFa$); driver presses brake pedal ($DrPu$); brakes malfunction ($BrMa$).

$$\begin{array}{lll} \mathcal{R} = & \{ & r_0 : BrFa \Rightarrow AcHa, \quad r_1 : \neg DrPu \Rightarrow AcHa, \\ & & r_2 : BrMa \Rightarrow BrFa, \quad r_3 : \neg DrPu \Rightarrow \neg r_2\}, \\ \mathcal{K} = & \{ & BrMa, \neg DrPu\}. \end{array}$$

The driver's omission to push the brakes ($\neg DrPu$) is a cause of the accident $AcHa$ (there exists a justified arguement for $AcHa$ based on $\neg DrPu$: $\neg DrPu \Rightarrow AcHa$). Brake malfunctioning $BrMa$ is not a cause: the brake pedal was not pushed so that the brake malfunction could not determine their failure, by argument $\neg DrPu \Rightarrow \neg r_2$.

War Crime (Necessary Causation or Ennoblement). In the following, we consider a war crime discussed in the literature [3] 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 soldier complies. We formalise it as follows: prisoner dies ($PrDi$); soldier/sergeant shoots prisoner ($SoSh/SeSh$).

$$\begin{array}{ll} \mathcal{R} & = \{r_0 : SoSh \Rightarrow PrDi,\ r_1 : SeSh \Rightarrow PrDi,\ r_2 : \neg SoSh \Rightarrow SeSh\}, \\ \mathcal{K} & = \{SoSh\} \end{array}$$

Shooting by the soldier $SoSh$ causes $PrDi$, according to argument $SoSh \Rightarrow PrDi$. Note $\neg SoSh$ would also cause $PrDi$. In a context in which the soldier does not shoot the prisoner (given $\mathcal{K} = \{\neg SoSh\}$), the following argument is justified: $[\neg SoSh \Rightarrow SeSh] \Rightarrow PrDi$ Therefore this example is an instance of ennoblement in a strict sense [34]: by shooting the soldier directly causes the death, and by not shooting he would indirectly cause the same outcome, inducing (ennobling) causation by the sergeant.

6 PREEMPTION WITH EVENT CALCULUS-STYLE MODELLINGS

We next formalise some of the above examples in a more refined way inspired by the event calculus, especially as used by Shanahan [27], to capture the temporal aspects of preemption scenarios in a more general way. In the approach of this section the defeasibility

[3] *Fairchild v Glenhaven Funeral Services Ltd* [2002] UKHL 22.
[4] *Dlugash v. People of State of NY*, 476 F. Supp. 921 (E.D.N.Y. 1979).
[5] The square brackets here indicate to which subargument r_4 is applied.
[6] Case 12124/2023, Cassazione Penale.
[7] *Saunders System Birmingham Co. v. Adams*, 61 A.L.R. 1333 (Ala. 1928).

of causal laws is, unlike in Section 5, not directly expressed as defeasible rules *cause* ⇒ *consequence*. Instead, causal laws are now 'reified' as first-order expressions of the form $\texttt{Causes}(a, f, t, t')$. In this notation variables $a, a', \ldots$ are used for actions (following [28], we use the term action in a generic sense, as a synonym of event), $f, f', \ldots$ for states of affair (called 'fluents' in the literature on the event calculus), and $t, t', \ldots$ for time points. To capture the defeasible nature of causal laws, we model them as consequents of defeasible rules of the form *Conditions* ⇒ $\texttt{Causes}(a, f, t, t')$ (where the condition part may be empty) .

The modelling below assumes discrete linear time plus a naming function for defeasible rules that for every rule r with free variables $x_1, \ldots, x_n$ returns the name $\texttt{Applicable}(r(x_1, \ldots, x_n))$. When the rule is instantiated, the variables are replaced by constants or function expressions with constants.

Rule r_1 addresses action-to-fluent causation: it says that if an action causes a fluent to happen, then the fluent holds after the action is executed. Rule r_2 addresses action-to-action causation: it says the same for when an action causes another actions to happen. Rule r_3 undercuts the action-to-fluent causation: it says that an action cannot cause a fluent to hold if the fluent already holds. This is a general way to model preemption for r_1. For r_2 and r_3 more specific undercutters must be given.

$r_1(a, f, t, t')$: $\texttt{Happens}(a, t), \texttt{Causes}(a, f, t, t') \Rightarrow \texttt{HoldsAt}(f, t')$

$r_2(a, a', t, t')$: $\texttt{Happens}(a, t), \texttt{Causes}(a, a', t, t') \Rightarrow \texttt{Happens}(a', t')$

$r_3(r_1(a, f, t, t'))$ $\texttt{HoldsAt}(f, t' - 1) \Rightarrow \neg\texttt{Applicable}(r_1(a, f, t, t'))$

The following rule scheme formalises temporal persistence: if a fluent holds at a certain point in time, it is presumed to hold in the future.

$r_4(f, t, t')$: $t < t', \texttt{HoldsAt}(f, t) \Rightarrow \texttt{HoldsAt}(f, t')$

Rule r_4 has the undercutter r_5: persistence of f from t to t' is no longer presumed if a new event causes the complement $-f$ of f at a t_2 between t and t'.

$r_5(r_4(f, t, t'), t_1, t_2)$: $\texttt{Happens}(a, t_1), \texttt{Causes}(a, -f, t_1, t_2), t < t_2 \leq t' \Rightarrow \neg\texttt{Applicable}(r_4(f, t, t'))$

This rule assumes a definition of the function symbol for negation, which can be formalised as a strict rule $\texttt{HoldsAt}(f, t) \rightarrow \neg\texttt{HoldsAt}(-f, t)$ and its transposition, which we leave implicit below in all examples.

6.1 Late Preemption: the Asynchronous Shooting

We now model the shooting case using the following predicates: $Tr(x, y)$, meaning that x tries to shoot y; $Sh(x, y)$, meaning that x shoots y; $Dead(y)$, meaning that y is dead. The knowledge base contains the facts that Bush tries to shoot Geller at time 1 and Dlugash at time 10.

$\mathcal{K} = \{\texttt{Happens}(Tr(Bu, Ge), 1), \texttt{Happens}(Tr(Dl, Ge), 10)\}$

$\mathcal{R} = \{c_1(x, y, t) :\Rightarrow \texttt{Causes}(Tr(x, y), Sh(x, y), t, t + 1),$
$c_2(x, y, t) :\Rightarrow \texttt{Causes}(Sh(x, y), Dead(y), t, t + 1)\}$

The knowledge base provides argument $B_1 = \texttt{Happens}(Tr(Bu, Ge), 1)$ while the instantiation of rule c_1 provides argument $B_2 = \Rightarrow_{c_1} \texttt{Causes}(Tr(Bu, Ge), Sh(Bu, Ge), 1, 2)$. By applying the appropriate instantiation of rule r_1, we obtain $B_3 = B_1, B_2 \Rightarrow_{r_2} \texttt{Happens}(Sh(Bu, Ge), 2)$. The corresponding instantiation of rule r_2 provides argument $B_4 = \Longrightarrow_{c_2} \texttt{Causes}(Sh(Bu, Ge), Dead(Ge), 2, 3)$. Through rule r_1, we get $B_5 = B_3, B_4 \Rightarrow_{r_1} \texttt{Holdsat}(Dead(Ge), 3)$.

Argument B_5 for $\texttt{HoldsAt}(Dead(Ge), 3)$ is justified, having no defeaters, and it includes subargument B_3 for $\texttt{Happens}(Sh(Bu, Ge), 2)$. Therefore, we can say that Bush's shooting is a cause of Geller being dead.

In a similar way, we can build arguments for Dlugash

$D_1 = \texttt{Happens}(Tr(Dl, Ge), 10)$
$D_2 = \Longrightarrow_{c_1} \texttt{Causes}(Tr(Dl, Ge), Sh(Dl, Ge), 10, 11)$
$D_3 = D_1, D_2 \Rightarrow_{r_2} \texttt{Happens}(Sh(Dl, Ge), 11)$
$D_4 = \Longrightarrow_{c_2} \texttt{Causes}(Sh(Dl, Ge), Dead(Ge), 11, 12)$
$D_5 = B_3, B_4 \Rightarrow_{r_1} \texttt{HoldsAt}(Dead(Ge), 12)$

with also D_5 concluding for the death of Geller.

However, D_5 is not justified (and thus Dlugash's shot is not a cause of Geller's death), since we can build a defeater of it. Indeed, by applying persistence rule r_4 we obtain an argument B_6 for $\texttt{Holdsat}(Dead(Ge), 11)$. Argument B_6 can be used to instantiate undercutter r_3, and thus obtain $B_7 = B_6 \Rightarrow -\texttt{Applicable}(r_1(Sh(Dl, Ge), Dead(Ge), 11, 12)$. B_7 defeats, and indeed overrules D_5, so that D_5 is no explanation of the death of Geller, and consequently the shot by Dlugash is no cause of it.

6.2 Symmetric Overdetermination: the Synchronous Shooting

We next illustrate with a minor modification of the above example that the above rules naturally deal with cases of overdetermination. Suppose now that Dlugash does not pull the trigger at time 10, so after Bush pulls his trigger, but at time 1, so at the same time as Bush. Then rules r_1 and r_2 are instantiated for Dlugash with the same time constants as for Bush, so the undercutter rule r_3 cannot be used, and the argument that Geller is shot dead by Dlugash is also undefeated. Then we have two explanations why Geller died, so both Bush and Dlugash caused Geller to die.

6.3 Early Preemption: the Non-Poisoned Bottle

The event-calculus based approach has the advantage of providing the tools to capture temporal inertia. This is suitable for modelling cases in which change happens through time, such as the following one (from [32]), where a full (non-empty bottle) is emptied.

> C is a traveller in the desert, whose only source of water is a keg full of water. A adds a fatal dose of undetectable poison to the water in the keg, for which there is no antidote. C remains unaware of the poison in the water. Subsequently, before C drinks any of the poisoned water, B dumps the poisoned water out of the keg. When C attempts to drink water from the keg, she discovers that it is empty. C dies due to dehydration.

Wright [32] argues that only B's emptying the keg is the cause of C's death, since the causal law for death by poisoning has as a condition 'poisoned water is in the keg', which is not satisfied.

The knowledge base and rule base are as follows.

$$\mathcal{K} = \{\texttt{HoldsAt}(-Empty, 0), \texttt{Happens}(Poisons, 1), \texttt{Happens}(Empties, 2), \texttt{Happens}(Thirst, 10)\}$$

$$\begin{aligned}\mathcal{R} = \{&c_1(t) :\Rightarrow \texttt{Causes}(Empties, Empty, t, t+1),\\ &c_2(t) :\Rightarrow \texttt{Causes}(Poisons, Poisoned, t, t+1),\\ &c_3(t) : \texttt{HoldsAt}(-Empty, t), \texttt{Holdsat}(Poisoned, t)\\ &\quad\Rightarrow \texttt{Causes}(Thirst, DrinksPoison, t, t+1),\\ &c_4(t) : \texttt{HoldsAt}(Empty, t)\\ &\quad\Rightarrow \texttt{Causes}(Thirst, Dehydratation, t, t+1),\\ &c_5(t) :\Rightarrow \texttt{Causes}(DrinksPoison, Dead, t, t+1),\\ &c_6(t) :\Rightarrow \texttt{Causes}(Dehydratation, Dead, t, t+1)\}\end{aligned}$$

We can build an argument for C's death by poisoning:

$P_1 = \texttt{HoldsAt}(-Empty, 0)$ (from $\mathcal{K}$)
$P_2 = \texttt{Happens}(Poisons, 1)$ (from $\mathcal{K}$)
$P_3 = \Rightarrow_{c_2} \texttt{Causes}(Poisons, Poisoned, 1, 2)$ (from $\mathcal{R}$)
$P_4 = P_2, P_3 \Rightarrow_{r_1} \texttt{HoldsAt}(Poisoned, 2)$ (fluent causation)
$P_5 = P_4 \Rightarrow_{r_4} \texttt{HoldsAt}(Poisoned, 10)$ (temporal persistence)
$P_6 = P1 \Rightarrow_{r_4} \texttt{HoldsAt}(-Empty, 10)$ (temporal persistence)
$P_7 = P_6, P_5 \Rightarrow_{c_3} \texttt{Causes}(Thirst, DrinksPoison, 10, 11)$ (from $\mathcal{R}$)
$P_8 = \texttt{Happens}(Thirst, 10)$ (from $\mathcal{K}$)
$P_9 = P_8, P_7 \Rightarrow_{r_2} \texttt{Happens}(DrinksPoison, 11)$
(from event causation)
$P_{10} = \Rightarrow_{c_5} \texttt{Causes}(DrinksPoison, Dead, 11, 12)$
$P_{11} = P_9, P_{10} \Rightarrow_{r_1} HoldsAt(Dead, 12)$ (from fluent causation)

Yet P_{11} is no explanation of the death (and thus the poisoning is no cause of it) since the following justified argument defeats P_{11}:

$D_1 = \texttt{Happens}(Empties, 2)$ (from $\mathcal{K}$)
$D_2 = \Rightarrow_{c_1} \texttt{Causes}(Empties, Empty, 2, 3)$ (from $\mathcal{R}$)
$D_3 = D_1, D_2 \Rightarrow_{r_4'} \neg\texttt{Applicable}(r_4(-Empty, 0, 10))$

Argument D_3 defeats argument P_{11} by undercutting its subargument P_6, i.e., by excluding that from the keg being non-empty at time 0 we can infer that it is still non-empty at time 10.

On the other hand we can build a justified argument for the death of C by using the fact that the keg is emptied at time 2, so that thirst causes dehydration and consequently death. Therefore, emptying the keg ($\texttt{Happens}(Empties, 2)$) can be considered a cause of C's death.

7 RELATED RESEARCH

7.1 Relevance in Argumentation

Above we already explained how our notion of an explanation improves on similar notions defined by [9] and [12]. Liao & van der Torre [18] propose and study principles for explanations of arguments in an abstract setting. We aim to study in future work to which extent our notions of an explanation satisfy their principles. For now we can already remark that we do not satisfy their assumption that all arguments have a unique explanation. Indeed, not satisfying this assumption is the main reason why we can apply our notions of relevance to NESS.

7.2 Causal Reasoning in Argumentation and AI & Law

The AI literature on causation is vast, so we cannot do more than briefly discuss some of the most important work, focusing on its connection to argumentation. The Halpern-Pearl account in terms of structural causal models [14] is seminal and takes some inspiration from NESS. However, their account essentially regards causal laws as deductive, with uncertainty modelled in probability distributions. One of our motivations for taking an argumentation approach is to model the defeasibility of reasoning with causal laws without using explicit probabilities. This fits the way in which causal generalisations are used in legal reasoning, where defeasible causal claims are often made and possibly attacked through counterarguments (for a defeasible cause-to-effect reasoning scheme see [29, Ch. 3], and for of defeasible causality in evidence, see [30]).

In [8] a detailed overview is given of applications of argumentation to causal reasoning, with as starting point the causal calculus of [7], which models causal rules as indefeasible. In their conclusion, the authors list the adaptation of [7]'s calculus with defeasible rules and exceptions as a topic for future research. As just noted, this was also a main motivation of our argumentation approach.

Of the work described in [8] that models actual causation, [4] too regards causal laws as indefeasible, since they encode structural causal models as preferred subtheories in assumption-based argumentation. The work of [5] and [6] models inference to the best explanation while [31] models its combination with causal prediction. Defeasible causal reasoning has also been studied within a logic programming approach (see recently [13]).

We know of no work on argumentation and causation that exploits the similarity between argumentation-based notions of relevance and NESS, although [9] note more generally that the notions of argumentation-based and causal relevance are similar.

Of other work in AI & law, the argumentation-scheme approach of [19] was a source of inspiration for our present approach. Lehmann & Gangemi [17] propose a formal ontology of concepts related to cause-in-fact, but do not model causal reasoning. Andreas et al. [1] aim to model overdetermination within a counterfactual approach to causation. However, they do not discuss how their approach relates to NESS (which was indeed motivated by the need to address overdetermination) but instead compare it to the Halpern-Pearl approach to actual causality [14]. Their formalism 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.

8 Conclusion

This paper has provided a novel formalisation, based on argumentation, of cause-in-fact in legal cases. Our formalisation includes a general approach to actual causation, which provides the central

aspect of causation in law. While this approach is inspired by leading approaches to actual causality [14] and the NESS framework of [32], it is distinctive in explicitly modelling defeasible causal rules and exceptions to them. Our causation-in-fact model successfully addresses various conundrums that have troubled legal analyses of causation, still bound to the idea of condition-sine-qua non (such as overdetermination and preemption).

The argumentation-based approach presented here can be further expanded in various directions, for instance by allowing argumentation frameworks with:

- Arguments for supporting or rejecting a causal rule, so that the acceptance of such rules can be justified through argumentation, according to the available knowledge and evidence:
- Arguments specifying the conditions for, or exceptions to, the legal relevance of causes-in-fact.

By relying on argumentation, our notion of a cause-in-fact can be integrated into more specific notions of legal causality (as proposed by doctrines of proximate or adequate causality) or with further conditions for legal responsibility (if we follow doctrines according to which the concept of causality should be restricted to causation-in fact). Moreover, by relying on argumentation the legal relevance or causes in fact may be excluded under general conditions (e.g., causation through omission is usually legally relevant only where there is an obligation to act), or blocked by legally specific exceptions (such as unforseeable actus novi leading to unexpected outcomes). While leaving this modelling to further work, we argue that an argumentation approach has the advantage of facilitating the integration of all such aspects within a unifying framework. A tool for the automatic evaluation of causation based on the presented formalisation is currently under development.

Finally, though we have here only focused on empirical causal connections (in order to specifically address causation-in-fact), our formalisation of a cause can have a broader significance, providing a concept of a reason or determinant of a defeasible conclusion, which may have applications in other domains, such as in normative and practical reasoning.

Acknowledgments

This research was funded by the European Research Council (ERC) Project "CompuLaw" (Grant Agreement No 833647) under the European Union's Horizon 2020 research and innovation program. "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).

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] P. Baroni, M. Caminada, and M. Giacomin. 2011. An introduction to argumentation semantics. *The Knowledge Engineering Review* 26 (2011), 365–410.
[3] S. Beckers. 2021. The counterfactual NESS definition of causation. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)*. 6210–6217.
[4] L. Bengel, L. Blümel, T. Rienstra, and M. Thimm. 2022. Argumentation-based Causal and Counterfactual Reasoning. In *1st International Workshop on Argumentation for eXplainable AI co-located with 9th International Conference on Computational Models of Argument (COMMA 2022), Cardiff, Wales, September 12, 2022 (CEUR Workshop Proceedings, Vol. 3209)*, K. Cyras, T. Kampik, O. Cocarascu, and A. Rago (Eds.). CEUR-WS.org. https://ceur-ws.org/Vol-3209/7343.pdf
[5] F.J. Bex, T. Bench-Capon, and K. Atkinson. 2009. Did he jump or was he pushed? Abductive practical reasoning. *Artificial Intelligence and Law* 17 (2009), 79–99.
[6] F.J. Bex, P. van Koppen, H. Prakken, and B. Verheij. 2010. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law* 18 (2010), 123–152.
[7] A. Bochman. 2021. *A Logical Theory of Causality*. MIT Press.
[8] A. Bochman, F. Cerutti, and T. Rienstra. 2024. Causation and argumentation. In *Handbook of Formal Argumentation*, D. Gabbay, G. Kern Isberner, G. Simari, and M. Thimm (Eds.). Vol. 3. College Publications, London, 627–715.
[9] A. Borg and F.J. Bex. 2024. Minimality, necessity and sufficiency for argumentation and explanation. *International Journal of Approximate Reasoning* 168 (2024), 109143.
[10] M. Caminada. 2014. Strong admissibility revisited. In *Computational Models of Argument. Proceedings of COMMA 2014*, S. Parsons, N. Oren, C. Reed, and F. Cerutti (Eds.). IOS Press, Amsterdam etc, 197–208.
[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] X. Fan and F. Toni. 2015. On computing explanations in argumentation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015)*. 1496–1502.
[13] Michael Gelfond and Evgenii Balai. 2020. Causal Analysis of Events Occurring in Trajectories of Dynamic Domains. In *ICLP Workshops*.
[14] J.Y. Halpern. 2016. *Actual Causality*. MIT.
[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] 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.
[18] B. Liao and L. van der Torre. 2020. Explanation semantics for abstract argumentation. In *Computational Models of Argument. Proceedings of COMMA 2020*, H. Prakken, S. Bistarelli, F. Santini, and C. Taticchi (Eds.). IOS Press, Amsterdam etc, 271–282.
[19] R. Liepiņa, A. Wyner, G. Sartor, and F. Lagioia. 2023. Argumentation schemes for legal presumption of causality. In *Proceedings of the 19th International Conference on Artificial Intelligence and Law*. 157–166.
[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. 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.
[22] M.S. Moore. 2024. Causation in the Law. In *The Stanford Encyclopedia of Philosophy* (Spring 2024 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University.
[23] 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.
[24] H. Prakken. 2005. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation* 15 (2005), 1009–1040.
[25] H. Prakken. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1 (2010), 93–124.
[26] H. Prakken. 2024. An abstract and structured account of dialectical argument strength. *Artificial Intelligence* 335 (2024), 104193.
[27] M. Shanahan. 1999. The Event Calculus explained. *Lecture Notes in Computer Science* 1600 (1999), 409–430.
[28] M. Shanahan. 2024. Talking about large language models. *Commun. ACM* 67 (2024), 68–79.
[29] D.N. Walton. 2005. *Argumentation Methods for Artificial Intelligence in Law*. Springer.
[30] R. Wieten, F. Bex, H. Prakken, and S. Renooij. 2018. Exploiting causality in constructing Bayesian network graphs from legal arguments. In *Legal Knowledge and Information Systems. JURIX 2018: The Thirty-first Annual Conference*, M. Palmirani (Ed.). IOS Press, Amsterdam etc., 151–160.
[31] R. Wieten, F. Bex, H. Prakken, and S. Renooij. 2022. Deductive and abductive argumentation based on information graphs. *Argument and Computation* 13 (2022), 49–91.
[32] R.W. Wright. 2011. The NESS account of natural causation: a reponse to criticisms. In *Perspectives on Causation*, R. Goldberg (Ed.). Hart Publishing, Chapter 14.
[33] R.W. Wright and I. Puppe. 2016. Causation: linguistic, philosophical, legal and economic. *Chi.-Kent L. Rev.* 91 (2016), 461.
[34] 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.