Computational Models of Legal Argument

Trevor Bench-Capon
University of Liverpool
tbc@liverpool.ac.uk

Katie Atkinson
University of Liverpool
katie@liverpool.ac.uk

Floris Bex
Utrecht University & Tilburg University
f.j.bex@uu.nl

Henry Prakken
Utrecht University
h.prakken@uu.nl

Bart Verheij
University of Groningen
bart.verheij@rug.nl

Abstract

Argumentation is central to law: in a legal dispute the opposing parties present their arguments, and the court determines which should be accepted. Consequently legal argumentation has been a prominent topic of research in AI and Law. In this chapter we will discuss the generation, evaluation and use of arguments in AI and Law. Our focus will be on the chronological development of techniques for these tasks.[1]

[1]During the finalization of this chapter, its initiator Trevor Bench-Capon passed away. He was a driving force behind its production, and added contributions until his final weeks.

1 Introduction

Argumentation is central to law. Consider for instance the following debate about the situation that Mary's bike is stolen and was bought by John:

A: Mary is the bike's owner.

B: Why?

A: She is the original owner.

B: I disagree. John is owner.

A: Why?

B: He is the buyer.

A: I disagree. He was not bona fide.

B: Why?

A: He bought the bike for €20.

B: I disagree. He bought the bike for €25.

A: You are right. That is still a reason he was not bona fide

In this brief argumentative dialogue, we already see several relevant phenomena. Initially a *conflict of opinions* is encountered, here about who is the owner of the bike. Also *claims* made are supported by *reasons*, here for instance about why there is ownership. Reasons can be *supporting or attacking*. For instance, Mary's original ownership supports her current ownership, and the fact that John bought the bike for €25 attacks that he paid €20. Furthermore, sometimes reasons are not about a claim, but about the *relation between a reason and a claim*. For instance, here John not being bona fide attacks the support relation between him being the buyer and being the owner. The example shows how argumentation can proceed in a *dialogue*, here between participants A and B. Apart from making claims and providing reasons, also *questions* are asked ('Why?').

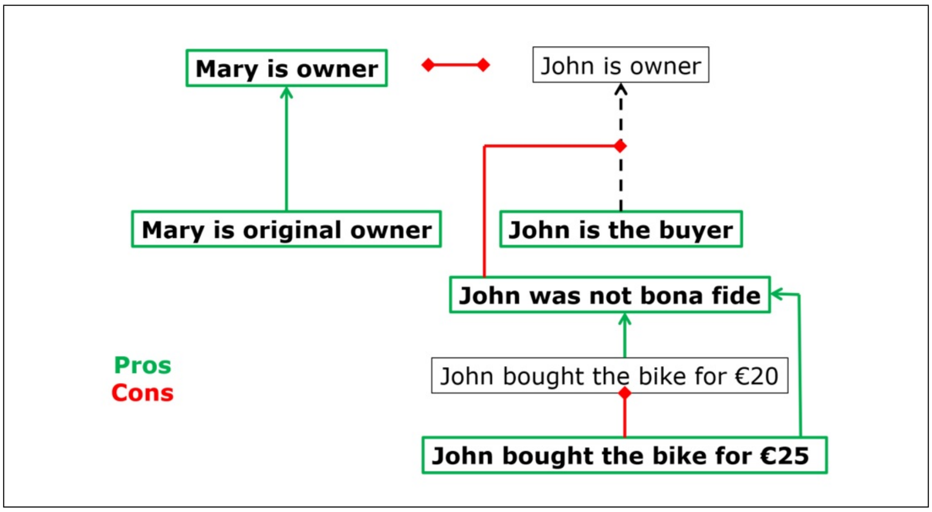

Figure 1: An example argument

In this dialogue, participants also make a *disagreement* explicit ('I disagree') which leads to a concession ('You are right').

Figure 1 provides a graphical representation of the arguments in the dialogue and their evaluation. Sentences in bold are accepted, either since they are undisputed claims (e.g. 'Mary is original owner') or since there is a successfully supporting reason for them (e.g. 'Mary is owner' supported by 'Mary is original owner'). Other sentences are not accepted, in fact they are rejected since there is a reason successfully attacking them (e.g. 'John is owner' attacked by 'Mary is owner'). Note that the figure shows that 'John was not bona fide' attacks the connection between him being the buyer and the owner. (A formalization is discussed in Section 5.3.4.)

Hence, since argumentation is so central, the topic of argumentation is prominent in research in AI and Law. For instance, the topics related to argumentation as they are discussed in the field of AI & Law include the following.

1. Legal cases have been studied from early on as the source of hypothetical arguments [Rissland and Ashley, 1987b, Ashley, 1990, Aleven and Ashley, 1995].

2. The dialogical use of legal rules, cases and values has been reconstructed as argumentation [Atkinson et al., 2005, Bench-Capon and Sartor, 2001, Gordon, 1993a, Hage et al., 1993, Loui and Norman, 1995, Prakken and Sartor, 1996, 1998].

3. Argumentation research has inspired schemes for decision-making and fact finding [Atkinson et al., 2005, Bench-Capon et al., 2000, Bex et al., 2003, Verheij, 2003c, Walton, 1996, Gordon et al., 2007].

4. Argument diagrams have been studied in the context of legal sense making [Bench-Capon, 1998, Bex et al., 2010, Gordon and Karacapilidis, 1997, Gordon et al., 2007, Verheij, 2003a].

5. Burden of proof has been analyzed in terms of argumentation [Gordon et al., 2007, Prakken and Sartor, 2007b, Prakken et al., 2005].

6. Legal decisions have been studied in early argument mining research [Mochales and Moens, 2011].

This correspondence between law and argumentation was also considered by philosophers, in particular Toulmin [Toulmin, 1958] and Perelman [Perelman and Olbrechts-Tyteca, 1969]. As a consequence, historically, AI and Law research has influenced computational argumentation research significantly, and vice versa. This in turn has led to the existence of various existing overview resources [Rissland et al., 2003, Reed and Norman, 2004, Bench-Capon et al., 2009, van Eemeren et al., 2014, Prakken and Sartor, 2015b, Bench-Capon, 2020, Verheij, 2020a, Atkinson and Bench-Capon, 2021].

In this chapter, we aim to add to these resources by giving a chronological presentation of the development of various techniques for computational argumentation in AI and Law, with a section devoted to each decade. We will organise our discussion around three generic tasks:

- Argument Generation
- Evaluation of Arguments
- Use of Arguments

2 Early days: Semi-formal approaches at the start of AI

In the 1950s, Toulmin [1958] suggested to radically change the analysis and assessment of reasoning in purely formal logic and probability theory by looking at the situated, concrete context of debate in law. Whereas his work was primarily philosophical in nature, work on argumentation in AI applied to the law started in the 1970s with McCarty's TAXMAN [1976], an early contribution to the reconstruction of legal argument in a formal-computational style. Rissland [1983] initiated the idea of hypothetical cases as examples guiding argument in the 1980s, taking inspiration from the use of examples in a mathematical discovery dialogue as suggested by Lakatos [1963].

2.1 Argument Generation

2.1.1 Prototypes and Deformations

Perhaps the first project to address argumentation in AI and Law was TAXMAN [McCarty, 1976]. In this project McCarty attempted to reconstruct the arguments in a number of leading US Tax Law cases. One particular case was *Eisner v Macomber* and in [McCarty and Sridharan, 1981] McCarty attempted to reconstruct the arguments of both the majority (justice Pitney) and dissenting (justice Brandeis) opinions, using the mechanism of *prototypes and deformations*.

The idea was that both start with a case representing a paradigmatic instance of their position (the prototype), and then map this into the current case through one or more mapping operations (the deformations). The issue in *Macomber* was whether payment of a dividend in the form of the distribution of additional shares in the same stock was taxable as income. The distribution of a corporation's cash, as in *Lynch v Hornby*, and the distribution of the stock of an unrelated corporation, as in *Peabody v. Eisner*, were situations that all parties agreed should be taxable. On the other hand, the appreciation in the value of a stock without the actual transfer of stock certificates, a purely hypothetical case, was a situation that all parties agreed should be non-taxable. Pitney's argument was the construction of a mapping between the Macomber case and the unrealized appreciation case: the taxpayer in Macomber now owns 3300 shares of common stock out of 75.000.000 outstanding, but, Pitney claimed, that is the same as owning 2200 shares out of 50.000.000 outstanding, which is the situation that would have existed had there been no actual transfer of stock certificates. A more difficult mapping to represent is the one constructed by Justice Brandeis to demonstrate that it is possible to find a coherent path between the stock distribution of *Eisner v. Macomber* and the cash distribution of *Lynch v. Hornby.* In his argument Justice Brandeis posits a sequence of hypothetical cases: first the distribution of common stock, then preferred stock, then bonds; then the distribution of long-term notes, then short-term notes; and finally the distribution of cash.

The mechanism is to represent cases as frames[2] and then starting from a precedent or a clear case (prototype), change various attributes (deformations) to map through a sequence of precedents and hypotheticals to reach the target case. Although there was not a full implementation of this process, since the search procedure to find a suitable sequence of mappings was not yet finalised, this was an important step in the computational modelling of legal argument.

2.1.2 Hypotheticals

Another early attempt to model legal reasoning was [Rissland, 1983]. Here the idea was to take a "seed case" and, by modifying various features of that case, generate a series of hypothetical cases to explore doctrines and approaches, and to uncover assumptions and biases. Although Rissland's original inspiration was mathematics [Lakatos, 1963] and [McCarty and Sridharan, 1981] is not given as a reference, there are similarities between this proposal and that of [McCarty and Sridharan, 1981]. If we take the current case as the 'seed case", we can see the process as an attempt to produce a series of hypotheticals leading to a prototype with the desired outcome.

[2]Frames [Minsky, 1975] were a then standard form of knowledge representation, in which entities were represented as frames, which contained a number of 'slots' corresponding to attributes of that kind of entity. Individuals were represented by instantiating the frame and filling the slots with the values of the attributes appropriate to that individual.

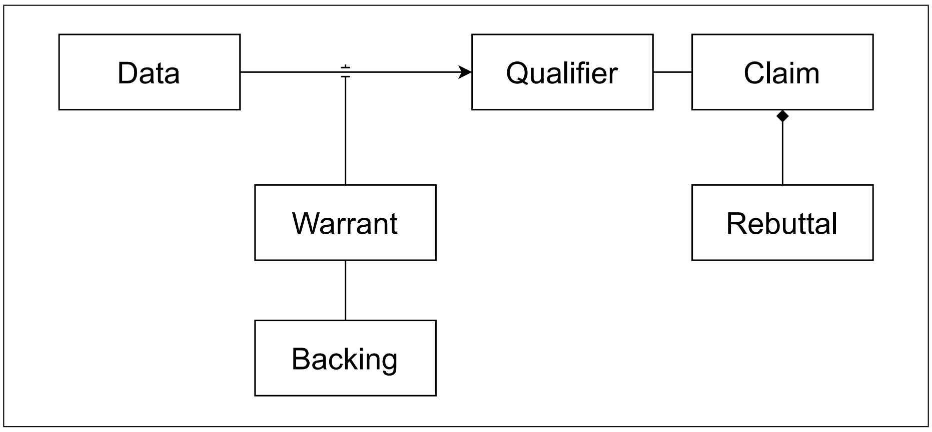

Figure 2: Toulmin's Argument Schema

Rissland's idea of using hypotheticals was more fully realised in the HYPO project with her then PhD student, Kevin Ashley, and we will give a more detailed discussion in Section 3.1.1.

2.2 Evaluation of Arguments

2.2.1 Toulmin

Toulmin was interested in encouraging critical thinking and as such in the defeasibility of most rules of inference. Very rarely does a set of premises entail its conclusion absolutely, in all circumstances. He therefore thought of argumentation in terms of justification rather than inference to a conclusion. His ideas are expressed in his argumentation schemes [Toulmin, 1958], shown in Figure 2.

Toulmin's scheme has six elements:

- **Claim**: This is the conclusion of the argument: note that Toulmin calls it a 'claim' to emphasise that the argument is intended to justify rather than establish it, and that it remains defeasible.
- **Warrant**: This is the rule used to justify the claim.
- **Backing**: This is the justification for the warrant. If the warrant is a legal rule, the warrant will be the statute or precedent case from which it derives.

- **Data** This is the basic premises needed to establish the antecedent of the rule.

- **Qualifier**: This expresses the degree of confidence in the claim, recognising that the warrants are rarely universally applicable, but often permit of exceptions. The qualifier will have different strengths depending on the nature of the warrant. Examples of qualifiers are 'certainly', 'probably', 'possibly', 'typically', 'usually'.

- **Rebuttal**: This represents exceptional circumstances under which the rule does not apply, expressed with an 'Unless' clause.

Toulmin's scheme thus adds the elements of Backing, Qualifier and Rebuttal to the standard *modus ponens* schemes of Premise (Data), Rule (Warrant) and Conclusion (Claim). This setup emphasises several ways in which an argument may be defeated: because the rule is unfounded in general, because it is inapplicable in the specific circumstances or because there is a stronger counter argument.

As we shall see later in the chapter, Toulmin's scheme had considerable influence in AI and Law, both for presentation (e.g. [Lutomski, 1989] and [Marshall, 1989]) and as a driver of dialogues [Bench-Capon, 1998]. This notion of an argumentation scheme was popular for a while, although it became replaced by Walton's more flexible notion of schemes [Walton, 1996].

Defeasibility was also an important aspect of the formalisms for argumentation developed in these early years. In recent structured accounts of argumentation (e.g. [Prakken, 2010]) three kinds of attack are identified. Two of these correspond to elements of Toulmin's scheme: the qualifier expresses that a warrant is defeasible, which allows for rebutting attacks on the claim, that is, arguments for a contradictory claim. Toulmin's rebuttals express explicit exceptions to warrants and are thus related to Pollock's undercutting counter arguments. Structured approaches to argumentation, however, do not require a backing for its rules: instead they allow undermining attacks, which are arguments claiming that the data is false, which did not arise for Toulmin, since there is no notion of chaining arguments in his scheme.

Toulmin's scheme underwent several adaptations by those interested in making it computable. This involved chaining schemes, so that the claim of one scheme became the data of another, leaving out the qualifier and/or the backing, redirecting the rebuttal to the rule rather than the claim, to represent undercut rather than rebuttal, and adding a Presumption element, justifying the rule by limiting the type of things to which it applied (e.g. [Bench-Capon, 1998]). Such adaptations will be discussed later in the context of particular systems which used them.

2.3 Use of Arguments

2.3.1 Logic Based Dialogue Games

[Hamblin, 1970] proposed an approach to the analysis of logical fallacies in terms of formal dialogues. The idea was that a dialogue would be formally specified as a set of rules, in such a way that the rules would prohibit the fallacy. The nature of the rules that would be broken if the fallacy is committed gives insight into the nature of the fallacy. These formal dialogue specifications became known in computational argumentation as 'dialogue games' (e.g. [Bench-Capon et al., 1991], [Gordon, 1993a], [Lodder and Herczog, 1995]).

[Hamblin, 1970] presented the dialogue game *H*, but it was Mackenzie's game *DC* [Mackenzie, 1979] that was the inspiration for several computational implementations including [Bench-Capon et al., 1991] and [Yuan et al., 2003]. The use of dialogue games both for providing interactive explanations, and as a means of modelling legal procedures became very popular in AI and law in the 90s, led by [Gordon, 1993a].

3 1980s: Rule-based and case-based knowledge representation

In the 1980s, adapting logical methods seems to be the way to go, since that is the language of computers. AI-wide this is the peak of nonmonotonic logic and logic programming [Gabbay et al., 1994]. Meanwhile, especially in the US, Case Based Reasoning (e.g. [Kolodner et al., 1985]) remained a widespread approach. In AI and Law, these developments give by the end of the decade prominent examples of a rule-based (British

Nationality Act, [Sergot et al., 1986]) and a case-based (HYPO, [Rissland and Ashley, 1987b]) approach.

3.1 Argument Generation

3.1.1 Dimensions in HYPO

In order to explore the generation and use of hypotheticals, identified as important in sections 2.1.1 and 2.1.2, Edwina Rissland and her PhD student, Kevin Ashley, conducted the HYPO project [Rissland and Ashley, 1987b, Ashley, 1989, 1990]. HYPO is perhaps the most influential project in AI and law and has inspired work in case based reasoning ever since [Bench-Capon, 2017]. HYPO contained several important ideas, but was very firm in its conception of case based reasoning as adversarial argumentation.

Three Ply Structure

In HYPO argumentation was modelled as a *three ply* activity, described in Section 3.3.1.

- **Citation**: In the first ply the proponent cites a precedent case with similarities to the current case and an outcome for the desired side.
- **Response**: In the second ply the opponent responds by citing a *counter example*, a precedent case with similarities by the opposite outcome, or by pointing to a *distinction*: a difference between the current case and the cited precedent which makes the current case stronger for the opponent; or by using a hypothetical to question one of the cited similarities.
- **Rebuttal**: in the third ply the proponent attempts to counter the argument of the opponent by distinguishing the counter examples. downplaying the distinctions and hypotheticals, and citing cases which show any weakness not to be fatal, or provide additional reasons for the desired outcome.

Similarity of Cases

In HYPO cases are represented as a collection of facts. These facts are then used to identify which dimensions are applicable to a case, and to assess the case in terms of these dimensions. A dimension is an aspect of a case which may have legal significance by presenting a reason to decide for one party or the other. The aspect takes a range of values, with one end representing the extreme pro-plaintiff value and the other the extreme pro-defendant value.

As originally conceived [Rissland et al., 1984], dimensions did not favour either party in particular, but could favour either depending on where on the range a particular case fell. An example would be the dimension of *SecurityMeasures*. At one extreme the plaintiff may have taken no security measures at all, which would be a reason to find for the defendant. At the other the extreme would be that the plaintiff had taken vert strict measures and this would be a reason to find for the plaintiff. In between which side is favoured is a matter for dispute, and courts will need to decide which, if any, side is favoured (moderate security measures may provide a reason for neither side). These decisions will become precedents, establishing how the dimension is used in future cases.

During the development of HYPO, however, Ashley's view of dimensions changed[3], and in [Ashley, 1990] he says that dimensions can be grouped "into those favoring the plaintiff generally and those favoring the defendant" ([Ashley, 1990], page 113)[4]. This shift was probably influenced by the nature of the dimensions implemented in HYPO. Ten of the thirteen are Boolean, with one value providing a reason, and the other not. For example, having a non disclosure agreement is a reason to find for the plaintiff, but the lack of one is not in itself a reason to find for the defendant, so it seems reasonable to describe this as a pro-plaintiff dimension. The three dimensions which do have a range, however, are less clear cut. *DisclosuresToOutsiders* was considered effectively binary since any disclosures at all were treated as a pro-defendant reason while no disclosures was not considered a reason for the plaintiff. With Security Measures, however, this breaks down: security measures is a dimension which can, and does, provide a reason for either side, and so cannot be classified as either pro-plaintiff or pro-defendant. The third

[3]For fuller discussions of the various different takes on dimensions and their evolution see [Rissland and Ashley, 2002] and [Bench-Capon and Rissland, 2001].

[4]By this time Ashley was already thinking in terms of factors, which would be come the basis of CATO [Aleven, 1997], as shown by [Ashley, 1991].

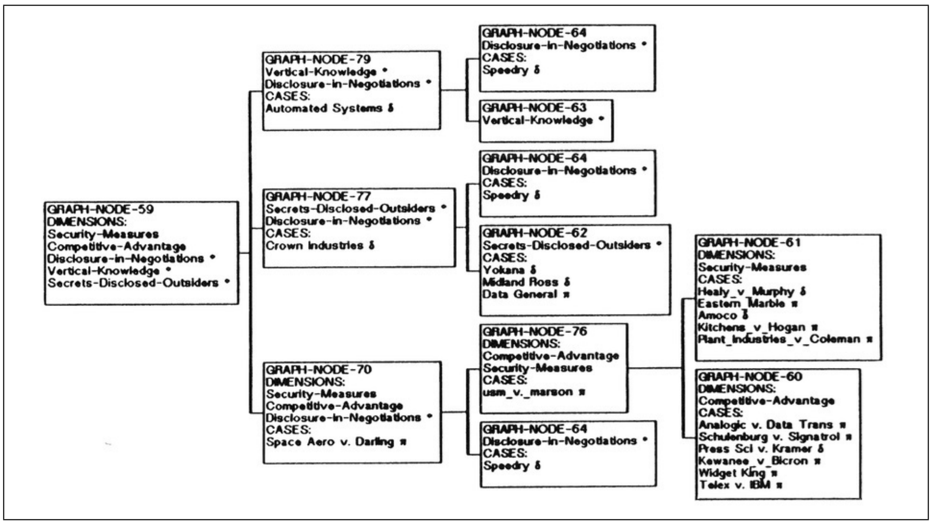

Figure 3: Claim Lattice used in HYPO taken from [Ashley, 1990]

numeric dimension, Competitive advantage, was considered by Ashley a pro-plaintiff dimension[5].

To determine the similarity between cases, the current case and the precedents are organised into a *claim lattice*. An example claim lattice, for *USM*[6] is shown in Figure 3[7]. The current case is shown as the root node, and its dimensions are listed. In Figure 3, *USM* has five dimensions. In the next level the nodes represent precedent cases with dimensions in common with the current case, where the dimensions in common are not a subset of any other precedent. For *USM*, there are three such cases, one with three dimensions in common and two with two dimensions in common. The next layer contains precedents with subsets

[5]This was criticised in [Bench-Capon and Gordon, 2022], where it was pointed out that the lack of competitive advantage could be seen as a reason to regard the information as not valuable and so to find for the defendant. This argument is in fact made in several precedent cases, whereas competitive advantage is rarely, if ever, given as a reason for the plaintiff, suggesting that the dimension is, if anything, pro-defendant.

[6]USM Corp. v Marson Fastener Corp. 379 Mass 90 (1979)

[7]Originally HYPO included "near miss" dimensions in the claim lattice [Rissland and Ashley, 1987b]. Rissland continued to use dimensions in their original sense and "near misses" play an important role in CABARET [Skalak and Rissland, 1992].

of these dimensions. Layers are added until we reach the leaves, which will have only a single dimension in common. Where precedents have the same dimensions in common they are represented in the same node. All the precedents in the lattice have some similarity to the current case: the closer to the root, the more similar they are.

Using the claim lattice we can construct the arguments to deploy in our three ply framework.

- We can cite a case closest to the root with the required outcome as a precedent. Thus in Figure 3, the plaintiff can cite *Space Aero*, and the defendant can cite either *Automated Systems* or *Crown Industries*.

- In the second ply, the respondent can cite a counterexample, such as a case supporting the respondent's side. Moreover it can distinguish the cited case by pointing to dimensions favouring the other side present in the root but not in the cited case, or to dimensions favouring the same side present in the cited case but not in the root, or to dimensions favoring the other side to a lesser degree in the root than in the precedent. Thus the defendant could respond to *Space Aero* by distinguishing with *Vertical-Knowledge* or *Secrets-Disclosed-Outsiders*. Finally, the defendant can also distinguish by saying that *SecurityMeasures* were less stringent or that the *CompetiveAdvantage* was less.

- In the third ply, counterexamples can be distinguished in the same way. The plaintiff could distinguish *Crown Industries* with *Security-Measures*. Distinctions can be rebutted by pointing to cases which also lacked the distinguishing feature: thus *Vertical-Knowledge* could be countered by pointing out that it was also not present in *USM* which was never the less found for the Plaintiff.

Thus the claim lattice can be used to generate arguments and counterexamples for both sides. The user is left to choose which arguments should be accepted.

Hypotheticals

Dimensions can also be used to generate hypothetical arguments, as discussed in [Rissland, 1989]. In the US Supreme Court, such arguments are typically used at the Oral Hearing stage to probe whether a particular dimension does indeed provide a reason to decide for the side mentioned. Here the idea is to consider a hypothetical case with a different point on the dimension. For example, suppose in Figure 3, the information had been disclosed to only seven outsiders, whereas in the precedent by the defendant, Crown Industries, the information had been disclosed to 150. Now one could argue that the current case is much weaker than *Crown Industries* on this dimension, and that in a hypothetical version of Crown Industries where it had been disclosed to only 50 outsiders, the plaintiff would have won. In this way Crown Industries is distinguished, since the current case is too weak on this dimension.

In [Ashley, 1990], four other ways of generating hypothetical cases are given (p. 148f.).

3.1.2 Logic Programs

In the 1980s the representation of legislation as logic programs was popularised with [Sergot et al., 1986]'s work on the British Nationality Act as a well-known example. Given such a logic program, it could be deployed as a legal expert system by adding a facility for the user to supply information as to the status of the leaf predicates. As an example consider US Trade Secrets Law[8] as discussed for HYPO in section 3.1.1. We have

```
TradesSecretsMisappropriation:- TradeSecret,
                                    Misappropriated.
TradeSecret:- InfoValuable, SecrecyMaintained.
Misappropriated if InfoUsed, Wrongdoing.
Wrongdoing:- BreachOfConfidence.
Wrongdoing:- IllegalMeans.
```

Such programs could explain their answer in the manner of the traditional *how?* explanations used in rule based systems since MYCIN [Buchanan and Shortliffe, 1984].

The problem is that the questions posed to the user are based on the terms of the legislation such as InfoValuable and BreachOfConfidence.

[8]This domain was not used by the original logic programmers, but we use it here to offer a direct comparison with HYPO.

But these terms are subject to interpretation, and need the clarification provide by case law. So reliable answers can only be given by a user expert in the case law of the domain.

To resolve this, the logic program was augmented with the reasons for applying these predicates established in case law. Thus, for example, that the information was disclosed in negotiations is a reason to find for the defendant, but that the defendant knew the information was confidential is a reason to find for the plaintiff. Thus we can add the clauses:

```
BreachofConfidence(d):-DisclosureInNegotiations.
BreachofConfidence(p):-KnewInfoConfidential.
```

Of course, both these things can be true in the same case, and so it is unclear in which way the issue should be resolved. In [Bench-Capon and Sergot, 1988] it was suggested that how? explanations of the set of answers could be seen as *arguments* for the two sides of the issue. Thus, here: *find for plaintiff since defendant knew the information was confidential* and *find for defendant since the plaintiff disclosed the information in negotiations.* The issue could then be resolved by choosing the better argument.

Generating arguments from a set of rules in this way became central to many current accounts of legal argumentation (e.g. [Prakken and Sartor, 1996, 1997], [Bench-Capon and Sartor, 2003], systems based on ASPIC+ [Prakken, 2010] and many more).

3.2 Evaluation of Arguments

3.2.1 Assessment by Users

In this period the emphasis was wholly on the generation of arguments. While both HYPO [Rissland, 1983] and the logic programming approach could generate arguments for both sides, they offered little support for choosing between them. The idea was that the users would evaluate the arguments on the basis of their knowledge and context.

Both systems were indeed often seen as being used before a trial by one of the parties to the case. In this scenario, the arguments for would be possible arguments to deploy in the trial, while the arguments against alerted the user to the potential counter arguments that might require rebuttal. In this scenario evaluation is unnecessary: the judge will have the ultimate decision.

Support for evaluation was left for future work. In [Bench-Capon and Sergot, 1988] the authors wrote:

> In the longer term, we hope to pursue what we have identified as a critical requirement: a representation in computer-intelligible terms of what it is that makes a 1egal argument persuasive.

Work on this was undertaken in the 1990s, as described in Section 4.2.

3.3 Use of Arguments

3.3.1 Three Ply

In HYPO [Rissland and Ashley, 1987a], arguments were deployed in the three ply structure described in Section 3.1.1. This three ply structure is common in Anglo Saxon law. The specific inspiration was the Oral Hearing stage of Supreme Court cases, in which the plaintiff makes a case, the defendant responds and the plaintiff rebuts, but there are other instances, such as witness examination in which, after the testimony has been elicited, there is a cross examination and a redirect.

The output from the program was a series of points relating to the three plies. Thus, using the claim lattice in Figure 3, we would get:

```
Point for Defendant as Side 1:
    Where plaintiff disclosed the information to outsiders
    and to the defendant in negotiations,
    defendant should win a claim for
                        TradeSecretsMisappropriation.
    Cite:  Automated Systems.

Response for Plaintiff as Side 2;
    Automated Systems is Distinguishable.
    The plaintiff took security measures.
    Not so in Automated Stystems.
```

```
    Space Aero provides a counter example in which
    Plaintiff took security measures,
    there was competitive advantage
    and plaintiff disclosed information in negotiations

Rebuttal for Defendant as Side 1;
    Space Aero is distinquishable
    The infomation concerned constitutes vertical knowledge
    Not so in Space Aero.
```

This structure is also represented in the argumentation schemes of [Walton, 1996], in which an instantiation of the scheme by a proponent is challenged by the opponent using characteristic critical questions, which the proponent must then attempt to answer. Indeed the reasoning of an immediate successor to HYPO, CATO[9] [Aleven, 1997], was modelled as a set of Walton style argumentation schemes in [Prakken et al., 2015].

3.3.2 How? and Why?

In the logic programming approach (e.g [Sergot et al., 1986]), the arguments were deployed using the explanation facilities commonly found in the expert systems of the time, modelled on [Buchanan and Shortliffe, 1984]. Thus when presented with a conclusion, the user could ask *how?* and be presented with the sequence of inferences which led from the entered facts to the conclusion. Thus, using the example program given in Section 3.1.2, suppose the user had said that the information was known to be confidential and the system had responded that there had been a trade secrets misappropriation. The *how?* query now yields:

```
I can show TradesSecretsMisappropriation
  because I can show  TradeSecret and Misappropriated.

I can show TradeSecret
   because I can show  InfoValuable and SecrecyMaintained.
```

[9] CATO is discussed in Section 4.1.5.

```
I can show Misappropriated
   because I can show InfoUsed and  Wrongdoing.

I can show Wrongdoing
   because I can show BreachOfConfidence.

I can show BreachOfConfidence
   because I can show KnowInformationConfidential.
```

Here InfoValuable, SecrecyMaintained and InfoUsed are all taken to default to the plaintiff: i.e. the burden of proof is on the defendant.

The *why?* explanation was used when the user was asked a question in the interaction. For example, the system might ask *Was the Information known to be confidential by the Defendant?*. If the user wants to know why this question is asked the why? query can be used, and will yield the following response.

```
If I know that the Information is Known Confidential
   I can show Wrongdoing.
```

Reiterated use of the *why?* query enabled the user to move up the proof tree and see why the goals were significant.

These two queries enable a fairly primitive dialectical dialogue between user and machine. This dialogical interaction underwent a great deal of development in the 1990s.

3.3.3 Toulmin presentation

Arguments can be presented as text, in dialogue, and also visually, in particular following the diagrammatic nature of the Toulmin argument scheme described in Section 2.2.1. An early example of diagrammatic presentation is provided by [Marshall, 1989], who made two adaptations to the original scheme, dropping the qualifier and allowing the chaining of arguments so that the claim of one argument became the data of the next. Basically this gave a visual presentation of the how? explanation described in Section 3.3.2, but with the useful addition of the backing for each step, which provided the source of the rules used. Additionally it was possible to provide a counter example using the rebuttal link. The

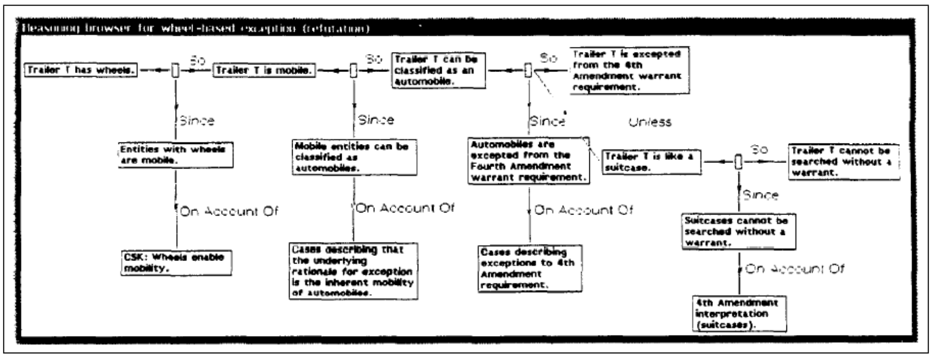

Figure 4: Marshall's presentation of an argument in *Carney*, taken from [Marshall, 1989]

nature of the exception was explained by making the rebuttal node the data of an argument with a contrary conclusion as claim. The approach was illustrated with the case of *California v Carney*[10], also used in [Rissland, 1989], and in subsequent AI and Law research on the automobile exception to the Fourth Amendment (e.g.[Bench-Capon and Prakken, 2010]).

Diagrammatic presentation of arguments was to become very popular both in computational argumentation in general (e.g. Araucaria [Reed and Rowe, 2004]) and in AI and Law in particular (e.g. Carneades [Gordon et al., 2007]).

4 1990s: Argumentation as an AI approach

In the 1990s, it became accepted that logic and logic programming do not suffice for the natural representation of debate. The focus turned to defeasibility, dialogue and procedure. Inspired by philosophy, argumentation takes center stage in AI ([Pollock, 1995, Dung, 1995]), and is immediately prominent in AI and Law. From the start, attempts are made to connect rules, cases, arguments in models of debate (in particular in the works of Bench-Capon, Prakken, Sartor, Hage, Gordon).

[10]California v. Carney :: 471 U.S. 386 (1985)

4.1 Argument Generation

4.1.1 Logic + Knowledge Base

That the *why?* explanation of logic programs could be seen as an argument comprising a series of *modus ponens* steps had been noted in [Bench-Capon and Sergot, 1988]. This idea was made more rigorous in [Prakken, 1993], where a formalisation of arguments and subarguments, conflicts between arguments and defeat was offered. The idea was that given a theory comprising facts and defeasible rules, arguments could be generated for and against a given statement. Particularly important was the idea of reinstatement, so that an argument which would otherwise be defeated by a preferred attacker could be reinstated if there was an argument to defeat that attacker.

This approach, generating arguments from an underlying knowledge base, was to become widespread, and is still used today in frameworks such as ASPIC+ [Prakken, 2010, Modgil and Prakken, 2014]. [Kowalski and Toni, 1996] advocated the use of assumption-based argumentation [Bondarenko et al., 1997] for the same purposes.

4.1.2 Rationales

Loui and Norman [1995] discuss rationales in legal decision making, addressing the formal explication of various kinds of argument moves that use rationales. We here follow the discussion by [Governatori et al., 2022] and [Bench-Capon and Verheij, 2022].

A key idea in the paper is that the rationales used in an argumentative dialogue can be interpreted as the summaries ('compilations') of extended rationales with more structure. By unpacking such summary rationales, new argument moves are possible. The paper distinguishes rationales for rules and rationales for decisions. In the authors' terminology, rule rationales express mechanisms for adopting a rule, while decision rationales express mechanisms for forming an opinion about the outcome of a case.

Here is an example of a small dialogue in which a compression rationale is unpacked, subsequently attacked and then defended against. The unpacking here has the form of adding an intermediate step, thereby interpreting a one step argument as a two step argument. We use a legal

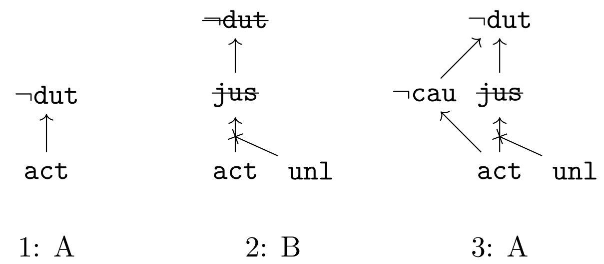

Figure 5: Unpacking a compression rationale

example (in the context of Dutch tort law), noting that the original paper focuses on abstract examples involving propositional constants a, b etc.

> A: I claim that there is no duty to pay the damages (¬`dut`) because of the act that resulted in damages (`act`).
> B: Unpacking your reasoning, you seem to claim ¬`dut` because of `act` using the additional intermediate reason that there is a ground of justification (`jus`). I disagree with `jus`, because the act was unlawful (`unl`), so there is no support for `jus`. Hence there is also no support for ¬`dut`.
> A: I agree with your reason `unl` and that hence there is no support for for `jus`. But I was not using `jus` as an intermediate step supporting ¬`dut`. Instead I used the intermediate step that there was no causal connection between the act and the damages (¬`cau`), hence my claim ¬`dut` because of `act`.

A graphical summary of the 3-step dialogue is shown in Figure 5. Normal arrows indicate a supporting reason and arrows ending in a cross indicate an attacking reason. All abbreviated statements are considered to be successfully supported, except those that are struck-through. Writing the first argument by A as `act` → ¬`dut`, B replies in the second move by interpreting the argument as actually having two steps `act` → `jus` → ¬`dut`, and then attacks the unpacked argument in the middle by the argument unl, making that jus and ¬dut are not successfully supported. But then at the third step A concedes that unl, while denying the unpacking via jus, instead claiming the unpacking act → ¬cau → ¬dut, providing an alternative way to support ¬dut, thereby still maintaining act → ¬dut.

4.1.3 Argument Moves

Deducing the consequences of a knowledge base provided a way of generating arguments for rule based approaches, but what of case based approaches, deriving from HYPO [Rissland, 1983]? Developments from HYPO took two distinct paths: Rissland worked with David Skalak on CABARET [Skalak and Rissland, 1992], while Ashley worked with Vincent Aleven on CATO [Aleven, 1997]. Both of them addressed argument generation through the use of *argument moves*.

4.1.4 CABARET

Arguments are generated in [Skalak and Rissland, 1992] with a three-tiered approach in terms of *argument strategies*, realised using *argument moves*, which are implemented using *argument primitives*. The appropriate strategy is selected by reference to the rule governing the case and the point of view. The move is determined by the precedents available and their dispositions. If the rule conditions are met and the point of view is positive, the hit must be confirmed but if the point of view is negative, the rule must be discredited. If the rule conditions are not met, the miss must be confirmed for a negative point of view, or the rule broadened for a positive point of view.

Once the strategy has been selected, the precedents are used to select a move. Depending on the outcome in the precedent and the strategy being employed the precedent must be analogised to or distinguished from the current case. These moves are then implemented through detailed comparison of the features of the current case and the precedent to determine the degree and nature of the matches and mismatches between the two[11]. For instance, when broadening a rule, citing a precedent with

[11] These primitives play the role of the factor partitions in [Wyner and Bench-Capon, 2007] and the functions in [Prakken et al., 2015] used in the instantiation of

the desired outcome that also failed to satisfy a rule antecedent, and so can be used to argue that since the missed condition was not necessary in that case, it is not needed in this case either.

In [Skalak and Rissland, 1992], the argument moves are limited to those which can be produced using the form of argument the authors term a *straightforward argument*, in which the facts of a current case are compared with a precedent case with the desired outcome. The paper, however, gives a taxonomy of argument forms used in legal argumentation, which includes a variety of additional forms of argument.

4.1.5 CATO

CATO [Aleven and Ashley, 1995, 1997, Aleven, 1997, 2003] was designed to help law students to distinguish cases effectively, and hence its emphasis was on distinguishing. The key point was that not every difference in the case could serve as a significant distinction. CATO replaced the dimensions of HYPO with *factors*. Factors are boolean and can be seen as ranges on dimensions favouring a particular party to the case and so providing a reason to decide for that party. Thus if a factor present in a precedent was absent from the new case, this would only provide a distinction if it favoured the winning side: if it favoured the losing side it would make the new case stronger than the precedent. But even so not all possible distinctions are considered significant: it may be that the difference does not weaken the case sufficiently to change the outcome.

To model this in CATO factors were organised into a *factor hierarchy*, (or rather five factor hierarchies, one for each issue). The issue would be at the root, with abstract factors coming between the issues and the base level factors (the factors corresponding to ranges on the dimensions) and serving to group them together according to whether their reasons was related. The children are reasons to think that their parent is present or absent. The factor hierarchy in CATO for the Issue of whether or not there was a confidential relationship is shown in Figure 6. Here we have two abstract factors (or 'intermediate legal concerns' as they are termed in [Aleven, 1997]), *NoticeOfConfidentiality* and *ExpressConfidentialityAgreement*, each of which is associated with

their argument schemes.

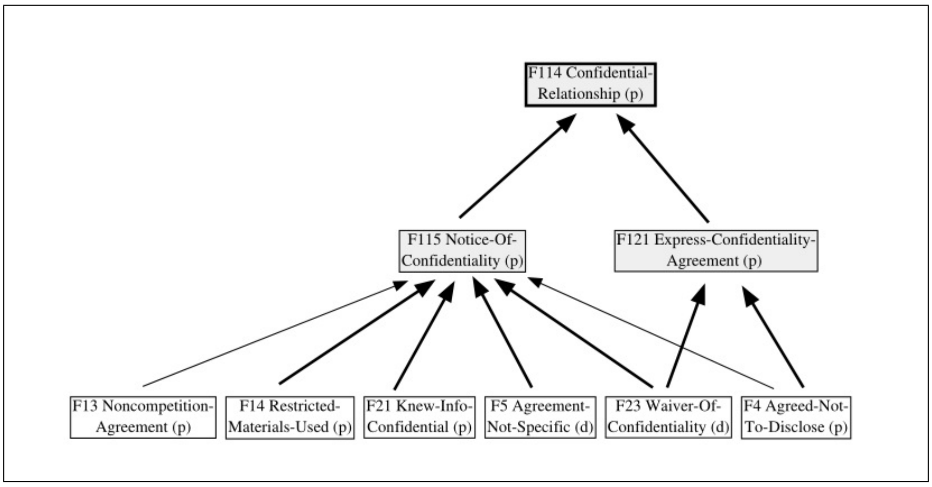

Figure 6: Factor Hierarchy for Confidential Relationship, taken from [Aleven, 1997]

a variety of base level factors, some favouring the plaintiff and some the defendant.

CATO uses the standard moves of citing a precedent, citing a counterexample and distinguishing a precedent, as found in HYPO and described in Section 3.1.1. But with the factor hierarchy, CATO can add additional moves to argue about the significance of a distinction.

Consider Figure 6. Suppose that we have a precedent with F14 Restricted Materials and F13 Non Competition Agreement. We might cite a new case which had F14 but lacked F13. If we did this our opponent could distinguish the case by pointing to the absence of F13. We may, however, there are other factors present in the cases which enable us to *downplay* the distinction, to argue that it is not significant.

This can be done in two ways

- If there is a factor with the same polarity in the current, we can argue that this factor can be substituted for this missing factor. For example if F21 Knew Information Confidential was in the new case;

- If there is a factor with a different polarity in the precedent case we can argue that that factor cancels the missing factor. For example, if F5 Agreement Not Specific had been in the precedent.

If, however, we can neither substitute nor cancel the distinguishing factor, our opponent can emphasise the significance of the distinction.

For a further discussion of CATO's argumentation moves see [Bench-Capon, 1997]. For a formal treatment of these moves in terms of argumentation schemes see [Wyner and Bench-Capon, 2007] and [Prakken et al., 2015].

4.1.6 Protoypes and Deformations

The idea of prototypes and deformations introduced in Section 2.1.1 was revived in [McCarty, 1995]. This paper claimed that knowledge representation languages available previously had been too inexpressive to implement the idea properly, in particular to formalise the notion of a prototype. Hence this paper presented an implementation, taking advantage of subsequent developments, in particular Language for Legal Discourse (LLD) [McCarty, 1989].

In [McCarty, 1995] we have a formalisation of the basic idea, illustrated with Prolog code, and a detailed computational reconstruction of the arguments of Justices Pitney and Brandeis in terms of the theory. But as noted in the discussion, while it was possible to generate the arguments, it was not possible to evaluate them: arguments that one was stronger than the other were not available. McCarty's suggestion was that this might in future be possible by considering the *coherence* of the competing arguments.

The notion of prototypes and deformation did not receive much subsequent take up, and rather CATO's factor based reasoning became the mainstream way of handling reasoning with precedent cases. What did, however, have significant influence was McCarty's notion of reasoning with precedents as *theory construction.* As he put it:

> Legal rules are not static, but dynamic. As they are applied to new situations, they are constantly modified to "fit" the new "facts". Thus the important process in legal reasoning is not theory application, but theory construction.

This idea was to prove influential in, for example, [Bench-Capon and Sartor, 2003] and [Chorley and Bench-Capon, 2005b], and was also the basis of [Prakken and Sartor, 1998], which used precedents to construct a theory comprising defeasible rules and priorities.

4.1.7 Heuristic Search

A rather different approach to argument generation was developed in BankXX. The system addressed the domain of bankruptcy (specifically the U.S. Bankruptcy Code, Chapter 13) [Rissland et al., 1996, 1997]. BankXX uses precedents (and other sources) to represent the domain knowledge as a highly interconnected network of building blocks, which is searched heuristically to gather *argument pieces*. The nodes in this network encompass a wide variety of ways of representing the domain knowledge, including cases as collections of facts, cases as dimensionally-analyzed fact situations, cases as bundles of citations, and cases as prototypical factual scripts, as well as legal theories represented in terms of domain dimensions. Thus cases are represented in several ways including, in its *Domain Factor Space*, "by a vector composed of the magnitudes of the case on each dimension that applies to it; non-applicable factors are encoded as NIL. This ... represents a case as a point in an n-dimensional space." Arguments are then formed by performing heuristic search over the network, using evaluation functions at the domain level, the argumentation piece level, and the overall argument level. The result is a highly sophisticated system, which provides a detailed analysis of the arguments available in a case. The approach is illustrated with a detailed case study of a particular case[12], and the system as a whole is subjected to a detailed evaluation in [Rissland et al., 1997].

The evaluation in [Rissland et al., 1997] is one of the most (if not the most) detailed examples of evaluation in AI and Law. It considers several different forms of the BankXX program, and the evaluation is conducted from several perspectives. A number of issues relating specifically to the evaluation of programs in the domain of law are noted.

BankXX has no obvious descendants in AI and Law research. Construction of cases by performing heuristic search was also carried out by AGATHA [Chorley and Bench-Capon, 2005a], but the search tree was over only a collection of cases represented as bundles of CATO-style

[12] In re Estus, 695 F.2d 311 (8th Cir. 1982)

factors, rather than the highly sophisticated network of knowledge used in BankXX.

4.1.8 Rule Based Representation of Precedents

Thus far we have seen how arguments can be generated from rule based representations using the proof trace of deductions from that rule base, and that arguments can be based on case based representations using the notion of similarity. The former had been primarily used for statute based reasoning and the latter for precedent based reasoning. The two were brought together in [Prakken and Sartor, 1998], which demonstrated a way of representing precedent cases and the decisions in these cases as a set of rules.

As was seen in Section 4.1.5, in CATO [Aleven, 1997], a case is associated with a set of factors, some pro-plaintiff and some pro-defendant, and an outcome. The pro-plaintiff factors offer reasons to find for the plaintiff and the pro-defendant factors offer reasons to find for the defendant. Now, if we have a decided case, C, containing pro-precedent factors P and pro-defendant factors D, then the conjunction of all factors in P will be the strongest[13] reason to decide C for the plaintiff and the conjunction of all factors in D the strongest reason to decide C for the defendant. The outcome in the case will show which of these two reasons is stronger. This means we have three rules:

r1 $P \rightarrow$ *plaintiff*;

r2 $D \rightarrow$ *defendant*;

r3 $C \rightarrow r2 \prec r1$ if the decision was for the plaintiff and $C \rightarrow r1 \prec r2$ if the decision was for the defendant.

[13] This assumes that the conjunction of two factors favouring the same side will always be stronger that the factors individually. This assumption is queried in [Prakken, 2005b], where an apparent counter example is given. However, such situations can be avoided by modelling the domain differently, using different factors for which the original “factors” are facts (e.g. [Horty and Bench-Capon, 2012], footnote 17). Arguably it is a necessary feature of factors as understood in [Aleven, 1997] they always favour a particular side, and this should hold whatever the context set by other factors [Bench-Capon, 2017].

This representation sees precedents as providing a one step argument from factors to outcome, which was the view taken in subsequent approaches such as [Bench-Capon, 1999] and the formalisations of precedential constraint stemming from [Horty, 2011]. In [Prakken and Sartor, 1998], however, Prakken and Sartor argue strongly that precedents should be seen in terms of multi-step arguments. Often the importance of a precedent will be with respect to a particular issue in the case [Branting, 1999]. Thus if we partition the factors according to the issues of the case, using, for example, the abstract factor hierarchy of [Aleven, 2003], we can get a finer grained representation of the argument. We now represent the case as $P_1 \cup D_1 \cup ... \cup P_n \cup D_n$, where P_i are the pro-plaintiff factors relating to issue i and D_i are the pro-defendant factors relating to issue i. We can now produce a set of three rules for each issue:

r4 $P_i \rightarrow I_i^P$, where I_i^P means that issue i is found for the plaintiff;

r5 $D_i \rightarrow I_i^D$; where I_i^D means that issue i is found for the defendant;

r6 $C \rightarrow r5 \prec r4$ if the issue was found for the plaintiff in C and $C \rightarrow r4 \prec r5$ if the issue was found for the defendant in case C.

We now write a set of three rules, using the issues in the antecedents to show how the issues determine the outcome. Suppose we have a case with three issues, of which two were found for the plaintiff and one for the defendant but the defendant won the case. This would give the rules:

r7 $I_1^P \wedge I_2^P \rightarrow \textit{Plaintiff}$

r8 $I_3^D \rightarrow \textit{Defendant}$

r9 $C \rightarrow r7 \prec r8$

Not only does this more faithfully reflect the reasoning in the case, but it has the practical advantage that an inference is not blocked by a distinction which is irrelevant because it pertains to a different issue. This two step reasoning, from factors to issues and then from issues to outcome was later used in IBP [Brüninghaus and Ashley, 2003] and Grabmair's VJAP [Grabmair, 2017]. More recently it has been argued that adopting this finer grained representation would improve the formal accounts of precedential constraint [Bench-Capon and Atkinson, 2021]. Even finer granularity would be possible, to give rise to three step arguments, but that will often associate too few factors with each sub-issue to be useful.

Using this representation we can generate arguments for both sides for a given issue, and also arguments based on precedents for which argument is the stronger.

4.2 Evaluation of Arguments

As we saw in Section 3.2, little had been done about the evaluation of arguments on the 1980s. In this decade, however, techniques for assessing competing arguments began to be developed.

4.2.1 Reason-Based Logic

In the 1990s, Hage developed Reason-based logic [Hage, 1993, 1996].[14] Hage presents Reason-based logic as an extension of first-order predicate logic in which reasons play a central role. Reasons are the result of the application of rules. Treating rules as individuals allows the expression of properties of rules. Whether a rule applies depends on the rule's conditions being satisfied, but also on possible other reasons for or against applying the rule. Consider, for instance, the rule that thieves are punishable:

$$\text{punishable: thief}(x) \Rightarrow \text{punishable}(x)$$

Here 'punishable' before the colon is the rule's name. When John is a thief (expressed as thief(john)), the rule's applicability can follow:

$$\text{Applicable}(\text{thief}(\text{john}) \Rightarrow \text{punishable}(\text{john}))$$

This gives a reason that the rule ought to be applied. If there are no reasons against the rule's application, this leads to the obligation to apply the rule. From this it will follow that John is punishable.

[14] Reason-based logic exists in a series of versions, some introduced in collaboration with Verheij (e.g. [Verheij, 1996]). The discussion here follows [van Eemeren and Verheij, 2018].

A characteristic aspect of Reason-based logic is that it models the weighing of reasons. In this system, there is no numerical mechanism for weighing; rather it can be explicitly represented that certain reasons for a conclusion outweigh the reasons against the conclusion. When there is no weighing information the conflict remains unresolved and no conclusion follows.

The formalization of Reason-based logic uses elements from classical logic and non-monotonic logic. Because of the emphasis on philosophical and legal considerations, the flavour of Reason-based logic is less that of formal logic, and comes closer to formally representing the actual ways of reasoning in the domain of law.

Reason-based logic has been applied, for instance, to a well-known distinction made by the legal theorist Dworkin [1978]: whereas legal rules seem to lead directly to their conclusion when they are applied, legal principles are not as direct, and merely give rise to a reason for their conclusion. Only a subsequent weighing of possibly competing reasons leads to a conclusion. Different models of the distinction between rules and principles in Reason-based logic have been proposed. Hage [1996] follows Dworkin and makes a strict formal distinction between rules and principles, whereas Verheij et al. [1998] show how the distinction can be softened by presenting a model in which rules and principles are the extremes of a spectrum.

4.2.2 Most Specific Argument

In Law, sometimes the following three principles are used to resolve conflicts between laws:

- *Lex superior*: prefer the law issued by the higher authority. Thus a national statute is preferred to a local by-law.

- *Lex specialis*: prefer the more specific law. Thus a law expressing an exception is preferred to the more general law.

- *Lex posterior*: Prefer the more recent law. Thus a new law overrules an existing law, and, in case law, the more recent decision is preferred.

Two of these, specificity and recency, had also been commonly used to resolve conflicts in Production Rule systems [Bench-Capon, 1990].

Inspired by [Poole, 1985], [Prakken, 1991] developed a formal theory based on preferring the most specific argument. In this paper, Prakken was mainly motivated by the need to handle exceptions to laws, which are indeed very common in law. The combination with the other two principles was addressed in [Prakken, 1993].

4.2.3 An Abstract Account of Argumentation

Later in the 90s, the world of computational argumentation was transformed by the introduction of Dung's notion of abstract argumentation frameworks [Dung, 1993a, 1995]. Dung's seminal idea was to represent a set of arguments and the attack relations between them, and then apply various semantics to identify acceptable sets of arguments. The key principle was that an argument is acceptable if and only if all its attackers are themselves attacked. Thus an argument may be defended by another argument which attacks its attacker, and these arguments may form an admissible set. To be *admissible*, a set must be conflict free (the members must not attack one another), and for all members of the set any attacker must be attacked by some member of the set. Two of the most important semantics are *preferred* semantics, which defines alternative sets of acceptable arguments as any *subset-maximal* admissible set, and *grounded* semantics, which defines a unique set of acceptable arguments as the least fixpoint of an operator that for any set of arguments returns the set of arguments defended by that set. There is always a single grounded extension, although it may be empty, but there may be multiple preferred extensions. In preferred semantics, therefore, sceptical acceptance, where an argument is in all preferred extensions, is distinguished from credulous acceptance, where an argument is in at least one preferred extension.

This abstract account of argumentation was taken up in AI and Law with papers by [Prakken and Sartor, 1996, 1997], [Kowalski and Toni, 1996] and [Jakobovits and Vermeir, 1999]. Prakken & Sartor defined their system for argument-based logic programming by defining the structure of arguments, an attack relation between arguments, and the use of priorities to determine which attacks result in defeats. By regarding the resulting defeat relation as the attack relation of abstract argumentation frameworks, this allows the use of abstract argumentation semantics for evaluating the arguments.[15]. This approach was later also applied in the ASPIC+ framework.

[Kowalski and Toni, 1996] developed a similar approach in the context of assumption-based argumentation. Instead of explicitly using priorities, they proposed to encode them in rule-exception structures.

[Jakobovits and Vermeir, 1999] defined various types of dialogue games to verify the acceptability of arguments in an abstract argumentation framework. They also studied dynamic games in which the argumentation framework can be extended during a dialogue, motivated by the observation that legal reasoning typically is an evolving process.

4.2.4 Burden of Proof

[Freeman and Farley, 1996] designed and implemented a model of legal argument in which arguments can be evaluated relative to a given level of proof. The language of their system divides rules into three epistemic categories: ‘sufficient’, ‘evidential’ and ‘default’, in decreasing order of priority. Arguments are structured as a variant of Toulmin’s argument structures (see Section 2.2.1 above) and can be of various types. Firstly, besides modus ponens the system also allows modus tollens. Moreover, it allows certain types of nondeductive arguments, viz. abductive ($p \Rightarrow q$ and q imply p) and a contrario arguments ($p \Rightarrow q$ and $\neg p$ imply $\neg q$). Taken by themselves these inferences clearly are the well-known fallacies of ‘affirming the consequent’ and ‘denying the antecedent’ but Freeman & Farley deal with this by also defining how such arguments can be attacked.

The strength of arguments is measured in terms of the four values ‘valid’, ‘strong’, ‘credible’ and ‘weak’, in decreasing order of priority. The strength depends both on the type of rule and on the type of argument. For instance, modus tollens results in a valid argument when applied to sufficient rules, but is a weak argument when applied to de-

[15] Strictly speaking Prakken & Sartor defined their system as an application of [Dung, 1993b] instead of [Dung, 1995], but their system can easily be recast as generating Dung’s abstract argumentation frameworks and applying grounded semantics to it

fault or evidential rules. Abduction and a contrario always result in just a weak argument. Finally, modus ponens yields a valid argument when applied to sufficient rules, a strong argument with default rules, and a credible argument with evidential rules. The strength of arguments is used to compare conflicting arguments, resulting in defeat relations among arguments, which in turn determine whether a move is allowed in a dispute.

Arguments can then be evaluated in terms of five different levels of proof, depending on which level is suitable in the given problem context:

- *scintilla of evidence* (find at least one defendable argument);
- *preponderance of the evidence* (find at least one defendable argument that outweighs the other side's rebutting arguments);
- *dialectical validity* (find at least one credible, defendable argument that defeats all of the other side's rebutting arguments);
- *beyond a reasonable doubt* (find at least one strong, defendable argument that defeats all of the other side's rebutting arguments);
- *beyond a doubt* (find at least one valid, defendable argument that defeats all of the other side's rebutting arguments).

Freeman & Farley motivate this approach by the observation that different legal problem solving contexts require different levels of proof. For instance, for the question whether a case can be brought before court, only a 'scintilla of evidence' is required, while for a decision in a case 'dialectical validity' is needed. Later, Tom Gordon incorporated these five levels of proof in his Carneades argumentation system [Gordon et al., 2007, Gordon and Walton, 2009] (see Section 5.1.4).

4.2.5 Social Values and Time Dependence

In their work on case-based argumentation in the law, Berman and Hafner emphasise the role of social values in the decision making of courts [Berman and Hafner, 1991, 1993, Hafner and Berman, 2002]. Such decision making is often purpose-oriented or teleological, in the sense that the purpose of promoting one social value may have to be balanced with the purpose of promoting another, competing value. Berman and Hafner write that legal precedents are 'embedded in a political context, where competing policies and values are balanced by the courts, and where legal doctrines evolve to accommodate new social and economic realities' [Hafner and Berman, 2002].

As an example of the balancing of social values, Hafner and Berman discuss cases about hunting wild animals. In one case, the plaintiff was a fisherman closing his large net, whereupon the defendant entered through the remaining opening and caught the fish inside (Young v Hitchens, 1844). Here there was a conflict between the competing social values of the pursuit of livelihood through productive work and economic competition. By deciding for the plaintiff or the defendant, a court can achieve the promotion of one value, but at the price of demoting the other. Here the court found for the defendant, but the judges' opinions show the careful balancing in the background. This case and the other wild animal cases have been extensively studied in Artificial Intelligence and Law, starting with [Bench-Capon, 2002a].

A specific theme addressed by Hafner and Berman is that the relevance of a case as an authoritative source to base new decisions on can evolve over time. The precedential value is not cast in stone, but develops over time influenced by societal changes. As their main example, they discuss a series of New York tort cases about car accidents. The issue was whether a driver should repair a passenger's damages. The series of cases are about what should be done when different jurisdictions are relevant, each with a different authoritative solution. For instance, when the driver and passenger are from New York, where the trip starts, and the accident happens in Ontario, Canada, should then the Ontario rule be followed—barring a law suit in such a case—or the New York rule where negligent driving could imply recovery of damages? Hafner and Berman discuss a series of cases that show the tension between a territory perspective, where the location of the accident (the *situs*) is leading, and a forum perspective, where the place of litigation determines the applicable law. Gradually, the cases shift from a strict territorial rule to a center-of-gravity rule, where the circumstances are weighed. Inspired by the work of Berman and Hafner, Verheij [2016a] developed a formalization of the example using techniques for the formal connection of qualitative and quantitative primitives (and the discussion here follows that paper). That formalization was used by Zheng et al. [2021] in an analysis of the hardness of case-based decisions and how that hardness changes over time.

4.3 Use of Arguments

4.3.1 Pleadings Game

In 1993 the idea of using nonmonotonic logics as a tool for formalising legal argument was already somewhat established. In [Gordon, 1993a,b] Tom Gordon added a new topic to the research agenda of the formalists in AI & Law: formalising the procedural context of legal argument. Gordon attempted to formalise a set of procedural norms for civil pleading by a combination of a nonmonotonic logic and a formal dialogue game for argumentation. The resulting Pleadings Game was not meant to formalise an existing legal procedure but to give a "normative model of pleading, founded on first principles", derived from Robert Alexy's [Alexy, 1978] discourse theory of legal argumentation.

The Pleadings Game had several sources of inspiration. Formally it was inspired by formal dialogue games for monotonic argumentation of e.g. [Mackenzie, 1979] and philosophically by the ideas of procedural justice and procedural rationality as expressed in e.g. [Alexy, 1978], [Rescher, 1977] and [Toulmin, 1958]. For example, Toulmin claimed that outside mathematics the validity of an argument does not depend on its syntactic form but on whether it can be defended in a rational dispute. The task for logicians is then to find procedural rules for rational dispute and they can find such rules by drawing analogies to legal procedures.

Besides a theoretical goal, Gordon also had the aim to lay the formal foundations for a new kind of advanced IT application for lawyers, namely, mediation systems, which support discussions about alternative theories by making sure that the rules of procedure are obeyed and by keeping track of the arguments exchanged and theories constructed.

The objective of the Pleadings Game is to support 'issue spotting', that is, to allow two human parties in a law suit to state the arguments and facts that they believe to be relevant, so that they can determine where they agree and where they disagree. The residual disagreements will go on to form the issues when the case is tried. The system plays two roles in this process: it acts as a referee to ensure that the proper procedure is followed, and records the facts and arguments that are presented and what points are disputed, so as to identify the issues that require resolution. The Pleadings Game has a built-in proof mechanism for an early argumentation-based nonmonotonic logic [Geffner and Pearl, 1992], which is applied to check the logical well-formedness of the arguments stated by the user, and to compute which of the stated arguments prevail, on the basis of the priority arguments also stated by the user and a built-in specificity checker. The Pleadings Game is truly dialogical since not only the content of the arguments is relevant but also the attitudes expressed towards the arguments and their premises.

Let us illustrate this with the following simplified dispute about whether a valid contract was concluded by the parties.

Plaintiff: I claim (1) we have a contract.
Defendant: I deny (1).
Plaintiff: We have a valid contract since (2) I made an offer and (3) you accepted it, so we have a contract.
Defendant: I concede (2) but I deny (3).
Plaintiff: (4) you said "I accept...", so you accepted my offer.
Defendant: I concede (4), but (5) my statement "I accept ..." was followed by terms that do not match the terms of your offer. This point takes priority (6) so I did not accept your offer.
Plaintiff: I concede the priority (6) but I deny (5).
Defendant: You required payment upon delivery (7) while I offered payment 30 days following delivery (8), so there is a mismatch.
Plaintiff: I concede (7) and the argument but I deny (8).

At this point, there is an argument for the conclusion that a contract was created using premises (2) and (4). The intermediate conclusion (3) of this argument that there was an acceptance is defeated by a counterargument using (7) and (8). So one outcome of the dispute is that no contract exists between the parties. However, in the Pleadings Game it also matters that the plaintiff has denied defendant's claim (8). This is a factual issue making the case hard, to be decided in court.

4.3.2 Other Dialogue Approaches

Other dialogue models of argumentation in AI and law have been proposed by Prakken and Sartor [1996, 1998], Hage et al. [1993], and Lodder [1999]. In Prakken and Sartor's approach (1996, 1998), dialogue models are presented as a kind of proof theory for their argumentation model (cf. Section 4.2.3). Prakken and Sartor interpret a proof as a dialogue between a proponent and opponent. An argument is justified when there is a winning strategy for the proponent of the argument. Hage et al. [1993], Lodder [1999] and Lodder and Herczog [1995] propose models of argumentation dialogues with the purpose of establishing the law in a concrete case. They are inspired by the idea of law as a pure procedure (though not endorsing it): when the law is purely procedural, there is no criterion for a good outcome of a legal procedure other than the procedure itself.

Some models emphasize that the rules of argumentative dialogue can themselves be the subject of debate. An actual example is a parliamentary discussion about the way in which legislation is to be discussed. In philosophy, Suber has taken the idea of self-amending games to its extreme by proposing the game of Nomic, in which the players can gradually change the rules.[16] Vreeswijk [1995] studied the game in a context of formal models of argumentative dialogues allowing self-amendments.

In an attempt to clarify how logic, defeasibility, dialogue and procedure are related, Prakken [1995] proposed to distinguish four layers of argumentation models. The first is the logical layer, which determines contradiction and support. The second layer is dialectical, which defines what counts as attack, counterargument, and also when an argument is defeated. The third layer is procedural and contains the rules constraining a dialogue, for instance, which moves parties can make, when parties can make a move, and when the dialogue is finished. The fourth and final layer is strategic. At this layer, one finds the strategies and heuristics used by a good, effective arguer.

Further dialog approaches from this period include [Bench-Capon et al., 2000, Bench-Capon and Staniford, 1995, Freeman and Farley, 1996, Bench-Capon, 1998].

[16] http://en.wikipedia.org/wiki/Nomic

4.3.3 Toulmin's Argument Model

Toulmin's argument model has been used in the context of information retrieval [Dick, 1991] and of the explanation of neural networks [Zeleznikow and Stranieri, 1995].

Dick [1991]'s starting point was that Boolean search could be enriched by the use of the conceptual structure underlying legal text. A proposal was made to analyze cases involving contract law using a frame-based representation of the elements of Toulmin's argument model. Case retrieval could then be achieved by matching frames.

Zeleznikow and Stranieri [1995] developed the Split-Up system in which knowledge-based modeling is combined with a neural network approach. The system addresses Australian family law, which by its discretionary nature cannot be fully represented in rule-based form. It is claimed that since neural network models can learn weights of relevant factors, they are well-suited for discretionary domains. However, in order to address the lack of explanations of decisions suggested by neural networks, a hybrid approach is developed in which the structure of Toulmin's argument model is used as an explanation format.

4.3.4 Argumentation and dialogue software

The theoretical developments on the modeling of argumentation and dialogue also led to various implemented software tools intended for support and guidance [Loui et al., 1997, Gordon and Karacapilidis, 1997, Verheij, 1999].

Room 5 [Loui et al., 1997] was intended as a testbed for public interactive legal argumentation. The user interface consisted of a web-based form that can be used to add reasons for and against claims in a public debate. The interface could list open cases, with also access to cases that are no longer argued. As an example, a local freedom of speech case was used. Argument structure is not visualized—as is more common—using boxes-and-arrows, but instead uses nested boxes ('encapsulated subargument frames'). Each box represents a claim, and a box in another box represents a reason relevant for a claim. Nested boxes to the left represent supporting reasons and to the right attacking. The representation format was developed to avoid the 'pointer spaghetti' that arises in a boxes-and-arrows format. The project's goals were not technical but to develop a web community of arguers trained in the use of the dialogue format, thereby building a database of semi-structured arguments.

The Zeno project [Gordon and Karacapilidis, 1997] was also meant to support and mediate online discussion. Its representation and interaction format combines elements of Toulmin's argument model with Rittel's Issue-Based Information System (IBIS). The approach includes issues, alternatives, positions (either for or against) and constraints that allow for the expression of preferences. The information entered by users is represented in a tree-like structure. Motivation for the specific approach in the Zeno project included the conceptual and computational complexity of the formalisms for nonmonotonic reasoning of that time.

ArguMed [Verheij, 1999] was intended as an argument-assistance system supporting a user's reasoning, to be distinguished from an automated reasoning system replacing the reasoning of a user. Statements entered could be assumptions or issues, for which supporting and attacking reasons could be given. It could be debated whether a reasoning step made was appropriate ('step warrants'; inspired by Toulmin's warrants) but also whether an attack of a reasoning step was appropriate ('undercutter warrants'). The system evaluated which statements were justified or not given the state of the discussion, graphically visualized using a boxes-and-arrows format. ArguMed was intended as a realization and testbed for theoretical argumentation models, and as step towards being a practical aid.

5 2000s: Deepening of the knowledge-data gap

In the decade following the year 2000, computational argumentation is becoming a field in itself, with its own conference series (Computational Models of Argument, COMMA),[17] still with significant influence from the field of law. Argumentation schemes and diagrams take off, and reasoning with values is formally analyzed. Burdens of proof are further studied, and evidence and fact finding in the law receive more attention. In the general field of AI, the gap between knowledge-based and data-driven methods is deepening. In connection with argumentation, there

[17] `http://www.computationalargumentation.org/`

is some work on prediction methods (see in particular Section 5.1.3), but this is not like the big data approaches that are arising. Knowledge-based approaches are on decline, while ontologies are an attempt to make a bridge with data-driven approacges. In the rest of the AI world, machine learning is gradually taking over, although rather neglected in AI and Law. Argument mining starts for real.

5.1 Argument Generation

5.1.1 Argumentation Schemes

In 2003, argumentation schemes were introduced in AI & Law by two articles in AI & Law journal [Bex et al., 2003, Verheij, 2003c]. Argumentation schemes are forms of argument that represent stereotypical patterns of human reasoning in a conditional form like rules. The idea of defining recurring patterns of reasoning through argumentation schemes originated with [Walton, 1996], who also studied it for legal and evidential reasoning ([Walton, 2002]). A well-known example of an argumentation scheme is the scheme for argument from expert opinion:

Argumentation scheme from expert opinion

Source e is an expert in domain d
e asserts that proposition a is known to be true (false)
a is within d

Therefore, a may plausibly be taken to be true (false)

In addition to a general rule or inference scheme, argumentation schemes also have associated critical questions that point to typical sources of doubt in an argument based on the scheme. For the scheme from expert opinion, the following six critical questions have been proposed ([Bex et al., 2003]):

1. Expertise Question: How credible is e as an expert source?

2. Field Question: Is e an expert in d?

3. Opinion Question: What did e assert that implies a?

4. Trustworthiness Question: Is e personally reliable as a source?

5. Consistency Question: Is a consistent with what other experts assert?

6. Backup Evidence Question: Is e's assertion based on evidence?

Answers to critical questions can lead to various types of counterarguments. For example, a negative answer to the "field question"would undercut an argument from expert opinion and a negative answer to the "consistency question" points to a possible rebutting counterargument with an opposite conclusion.

The idea of argumentation schemes is very closely related to Toulmin's notion of warrants (Section 2.2.1) and logical rules (Section 3.1.2). In fact, [Verheij, 2003c] argued that argumentation schemes should be used as the basis for a *logic of law* that focuses on the specific domain rules and contextual reasoning patterns in law. Both he and [Bex et al., 2003] formalised the argumentation schemes and their critical questions in logics for structured argumentation (namely [Verheij, 2003b] and [Prakken, 1993], respectively). Note that not all such schemes are schemes for reasoning with evidence and facts. For example, [Verheij, 2003c] also provides an example of a more legal scheme:

Person p has committed crime c
Crime c is punishable by n years of imprisonment

Therefore, person p can be punished with up to n years in prison

5.1.2 Stories and explanations

In addition to the rule-based and case-based approaches to (legal) argumentation, the 2000's also saw the rise of *story-based* argumentation, mainly in reasoning with evidence (see Section 5.3.3). The story-based approach to reasoning stems from legal psychology ([Pennington and Hastie, 1993, Wagenaar et al., 1993]), and focuses on stories about what happened in a legal case, that is, the facts in the case. These hypothetical stories, coherent sequences of events connected by (sometimes implicit) causal links of the form c *is a cause for* e, are used to explain the observed evidence in a case. When explaining some observed event e, we perform what is commonly called *causal–abductive reasoning* ([Josephson and Josephson, 1996]): If we have a general causal rule

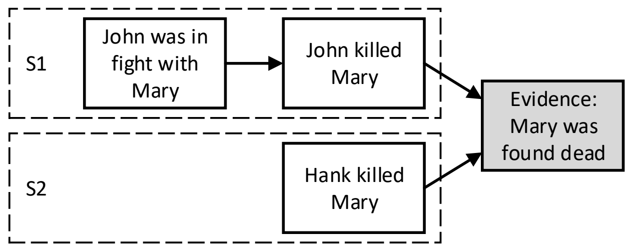

Figure 7: Different stories explaining the evidence, where the arrows indicate causal links.

$c \rightarrow e$ and some observed evidence e, we can infer cause c as a possible explanation of effect e. This cause can be a single event, but it can also be a sequence of events, a story. Taken by itself the abductive scheme is nothing but the fallacy of affirming the consequent. However, in a setting where alternative abductive explanations are generated and compared, it can still be rational to accept an explanation if no better other explanation is available.

Like argumentation, reasoning with stories is dialectical, in that different competing explanatory stories are compared. Take, as an example, the two stories in Figure 7, where two possible explanations for the observation that *Mary was found dead* are provided: one story where John killed Mary, and another one where Hank killed Mary. These two stories are alternative explanations for the evidence, and we have to choose between them by, for example, looking for new evidence that supports or attacks the different stories. This reasoning with, and about, stories was first formalised by [Keppens and Zeleznikow, 2003], and later in a series of articles by Bex, Prakken and Verheij ([Bex et al., 2006, 2007a]), in which the *hybrid theory* of stories and arguments is proposed, where individual arguments based on evidence are used to support and attack the different hypothetical stories in the process of inference to the best explanation. This will be further discussed in Section 5.3.3 on evidence.

In addition to stories explaining the evidence as in Figure 7, it is also important that a story is *plausible*: irrespective of the evidence, does the story fit with our ideas about how things generally happen? Plausibility plays a big part in our reasoning. For example, we would not seriously consider the scenario 'Aliens killed Mary' because this is highly implausible. Furthermore, elements which are implausible at first sight might warrant further investigation: for example, if John has no history of violent behaviour, it seems implausible that he would immediately kill Mary after getting into a (verbal) fight with her (i.e. the causal link between *John was in a fight with Mary* and *John killed Mary* is implausible). Furthermore, stories can contain gaps, missing elements that make them less plausible. One way to look for such gaps is to compare the story to story schemes ([Pennington and Hastie, 1993, Bex, 2009]) or scripts ([Schank and Abelson, 1977]), stereotypical patterns that serve as a scheme for particular stories. Take, for example, a general scenario scheme for intentional action: a motive leads to an action, which has certain consequences. In our example, S2 is less plausible than S1 because it does not include a motive for why Hank killed Mary.

Such reasoning about motives in stories was the subject of further work by Bex, Atkinson and Bench-Capon ([Bex et al., 2009], who define an argumentation scheme for *abductive practical reasoning* based on the regular (non-abductive) scheme for practical reasoning (Section 5.2.2).

Argumentation scheme for abductive practical reasoning

The current circumstances C
are explained by the performance of action A
in the previous circumstances R
with motivation M

Possible critical questions for this scheme are, for example, *Are there alternative ways of explaining the current circumstances S?* or *Can the current explanation be induced by some other motivation?*. Following [Atkinson and Bench-Capon, 2007], the abductive scheme and its critical questions were formalised as action based alternating transition systems, providing a formal semantics for abductive reasoning about motives using stories.

5.1.3 Issue-Based Prediction

[Brüninghaus and Ashley, 2003] and [Ashley and Brüninghaus, 2009] proposed a system for Issue-Based Prediction (IBP) as a descendant of the HYPO and CATO systems (see Sections 3.1.1 and 4.1.5). It predicts outcomes of US trade secret misappropriation cases and provides explanations for its predictions in terms of an argumentation model that combines rule- and case-based reasoning. Cases are as in CATO represented as two sets of factors favouring, respectively, the plaintiff and the defendant. IBP's knowledge model combines a logical decision tree with lists of pro-plaintiff and pro-defendant factors for each of the five leaves of the tree, called the issues (e.g. *did the plaintiff maintain secrecy*, and *did the defendant obtain the secret by improper means?*), as shown in Figure 8. Issues are addressed with a prediction model that according to Ashley and Brüninghaus applies a kind of scientific evidential reasoning. Roughly, if all factors in the case favour the same side for that issue, then IBP predicts a win for that side on the issue (unless all these factors are 'weak'). Otherwise it retrieves precedents that contain all case factors on that issue. If all have the same outcome, then IBP predicts that outcome, otherwise it tests the hypothesis that the side that won the majority of precedents will win, by trying to explain away each precedent won by the other side; this attempt succeeds if the precedent contains a 'knock-out' factor that is not in the current case. IBP's notions of weak and knock-out factors are a refinement of the CATO factor model and are defined in terms of low, respectively, high predictive power for the side they favour. Finally, IBP's predictions on all the issues are combined in an overall prediction.

In an evaluation experiment IBP outperformed 11 other outcome predictors and achieved a high accuracy score of 92%. Although, strictly speaking, IBP does not reason about what to *decide* but about what to *predict* as a decision, [Ashley, 2019], quoting [Aleven, 2003], claims that predictive accuracy is a good (although not the only) measure of the reasonableness of a computational model of argument.

5.1.4 Carneades

In [Gordon et al., 2007, Gordon and Walton, 2009] Tom Gordon and co-authors proposed a new formal argumentation system, with various sources of inspiration. One was [Freeman and Farley, 1996]'s model of argument evaluation with five alternative levels of proof, and another was [Walton, 1996]'s dialogical theory of argumentation schemes. Like Gor-

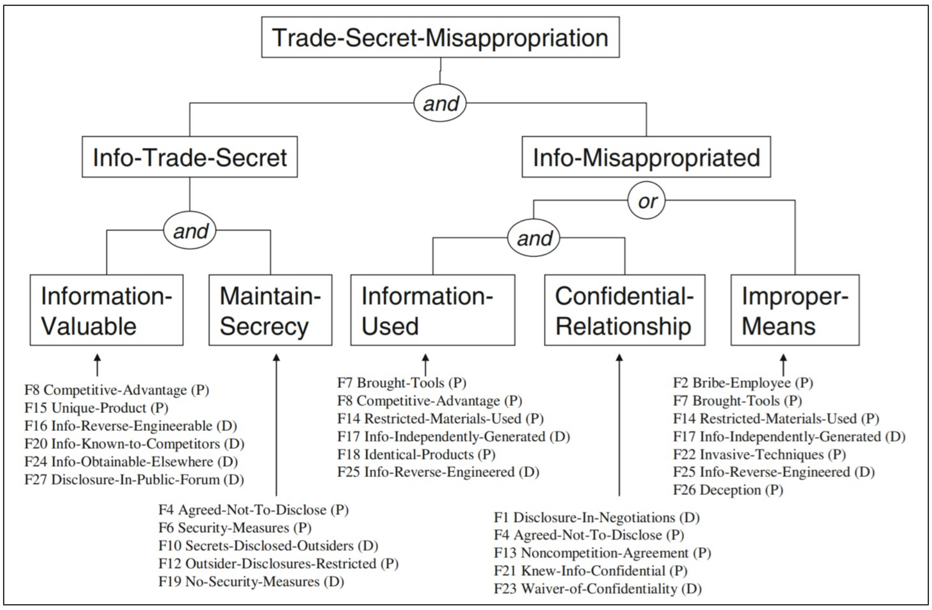

Figure 8: Hierarchy of Issues and Factors, taken from [Ashley and Brüninghaus, 2009]

don's earlier Pleadings Game (see Section 4.3.1) the Carneades system is meant to be used in a dialogical context, although unlike the Pleadings Game it does not explicitly generate dialogues but only records which statements have been accepted or rejected by a given audience. It then incorporates this information in its evaluation of arguments and statements.

Unlike, for instance, ASPIC+, Carneades does not evaluate arguments by generating Dung-style abstract argumentation frameworks. Instead, in Carneades each statement can be assigned its own standard of proof. The system takes not proof burdens but proof standards as the primary concept, and encodes proof burdens with particular assignments of proof standards. A Carneades argument has a set of premises P, a set of exceptions E and a conclusion c, which is either pro or con a statement. Carneades does not assume that premises and conclusions are connected by inference rules but it does allow that arguments instantiate argument schemes. Also, all arguments are elementary, that is, they contain a single inference step; they are combined in recursive definitions of *applicability* of an argument and *acceptability* of its conclusion. In essence, an *argument* is *applicable* if (1) all its premises are given as a fact or are else an acceptable conclusion of another argument and (2), none of its exceptions is given as a fact or is an acceptable conclusion of another argument. A *statement* is *acceptable* if it satisfies its proof standard. Facts are stated by an *audience*, which also provides numerical *weights* of each argument plus *thresholds* for argument weights and differences in argument weights. Three of Carneades' proof standards are then defined as follows:

> Statement p satisfies:
>
> - *preponderance of the evidence* iff there exists at least one applicable argument pro p for which the weight is greater than the weight of any applicable argument con p.
> - *clear-and-convincing evidence* iff there is an applicable argument A pro p for which:
>   ∗ p satisfies *preponderance of the evidence* because of A; and
>   ∗ the weight for A exceeds the threshold α, and
>   ∗ the difference between the weight of A and the maximum weight of the applicable con arguments exceeds the threshold β.
> - *beyond-reasonable-doubt* if and only if p satisfies *clear-and-convincing evidence* and the maximum weight of the applicable con arguments is less than the threshold γ.

Although Carneades was not set up with the aim to generate Dung-style abstract argumentation frameworks, [Van Gijzel and Prakken, 2012] translated Carneades into Dung's frameworks via the ASPIC+ framework of [Prakken, 2010], showing that Carneades induces a unique extension in all semantics. [Brewka and Gordon, 2010] give an alternative reconstruction of Carneades in terms of [Brewka and Woltran, 2010]'s abstract dialectical frameworks. In [Gordon, 2013] a web-based implementation of Carneades is described, including an argument visualisation tool. Some attempts have been made to connect Carnaedes to data by combining ontologies with argumentation [Gordon, 2011].

5.1.5 A Rule-Based Approach: Defeasible Logic

Governatori and others developed an approach to legal knowledge representation and reasoning in the context of Defeasible Logic [Nute, 1994]. This logic does not have an explicit notion of an argument but models recursive notions of defeasible and strict derivability in terms of the application of possibly conflicting prioritized rules. Governatori and colleagues paid much attention to various aspects of legal knowledge representation, such as deontic notions, time and change [Governatori et al., 2005, 2007, Rotolo and Governatori, 2009].

5.1.6 Argumentation Mining

The idea of argumentation mining first arose in AI & Law: [Grover et al., 2003, Hachey and Grover, 2005, 2006] developed machine learning models for the automatic detection of pieces of text that represent rhetorical roles in UK House of Lords judgements. Such rhetorical roles are the elements of arguments in legal texts, for example, facts, citations, but also more direct argumentative types of roles such as refutations against or argumentations for claims. The classification results (i.e. what rhetorical role does this sentence play?) were quite good for a 7-class problem, with F-scores around 55-60%.

Argumentation mining as a separate task (as opposed to a sub-task of summarization as with [Hachey and Grover, 2006]) was widely popularised by the seminal work of [Moens et al., 2007] and [Palau and Moens, 2009] (which was later extended and published as [Mochales and Moens, 2011]) . Where [Moens et al., 2007] detected just elements (i.e. premises, (sub)conclusions) of arguments, [Mochales and Moens, 2011] also detected the structure of arguments, that is, the inference relations between the premises and conclusions. Such detection of structures was generally more difficult than element detection: an accuracy of 60% was achieved when detecting argumentation, while an F1-score of 70% was achieved for recognizing premises and conclusions.

5.1.7 Machine Learning

During this decade machine learning approaches were rather neglected. There was little interest in neural networks as used in the 1990s (e.g. [Bench-Capon, 1993] and [Zeleznikow and Stranieri, 1995]) and the availability of large datasets and techniques for learning from them had not yet reached the stage where they could support later big data approaches [Villata et al., 2022]. Still the decade did produce two interesting examples of using machine learning for argumentation.

The first example concerns Argument Based CN2 (ABCN2), [Možina et al., 2005], is an extension of the well-known rule-learning algorithm CN2 of [Clark and Niblett, 1989]. CN2 is an inductive logic programming algorithm that produces a set of rules that can be used to classify instances in the domain. The central idea is to augment CN2 to accept, along with data, arguments explaining the classification of a small number of instances to improve both the efficiency of the learning process, and the quality of the rules learned. The arguments constrain the search space, and, it is claimed, induce rules that make more sense to domain experts. The process is iterative. After the first pass, the most frequently misclassified example is presented to the expert to give an argument as to why it should be classified in the correct way. The second pass then begins from a rule induced from the expert's argument. The process continues until there are no problematic examples.

The example study was the fictional welfare benefit data first used in [Bench-Capon, 1993] and later used in other projects and now publicly available [Steging et al., 2023]. This example took seven iterations. The rules induced were close to the ideal set, except that two thresholds were slightly low: 59 rather than 60 and 2900 instead of 3000, because there were no examples in the dataset to identify the thresholds precisely.

Further experiments tested robustness in the face of incorrect data. This is important since we cannot guarantee that all examples will have been correctly classified, especially in a domain like welfare benefits, where the error rate is notoriously high. Various noise levels were tested up to 40%. The results showed that ABCN2 outperformed the original CN2 at every level, with the gap widening as noise exceeded 10%.

A second example of the use of Machine Learning in the 2000s was [Ashley and Brüninghaus, 2009]. This paper describes the augmentation of IBP (See Section 5.1.3) with a program, SMILE, which will ascribe factors given natural language input. This means that there is no need to manually analyse the cases: together the programs can predict an outcome based on a natural language description of the facts. SMILE used a dedicated set of rules for each factor. There is no need to dwell on details here, since natural language techniques are now vastly better.

The performance using SMILE fell off drastically from that which had been achieved using IBP with manually ascribed factors outcome-prediction accuracy dropped from 92% to 64%. This suggests that the learning to ascribe factors is rather hard. A better comparison is with machine learning approaches such as [Medvedeva et al., 2019]. Such approaches also fall well short of 90% accuracy, typically achieving something in the 70-85% range. This is true for [Medvedeva et al., 2019], which was tested in the domain of the European Convention on Human Rights, although it fell to 52% for Article 10 with the best performance of 84% on Article 16. Average performance across all articles was 74%. It should be noted that these recent approaches do not learn to extract factors from case texts but instead immediately relate the natural-language case texts to outcomes.

Since the logical model can be constructed to a high degree of accuracy - and, importantly, can provide arguments to justify the prediction - some argue that the machine learning should be used for factor ascription and the outcome determined with a logical model (e.g. [Mumford et al., 2022]). Whether, however, sufficiently accurate performance in ascribing factors can be achieved has yet to be shown. An alternative approach in [Prakken and Ratsma, 2022] is to use a logical model to provide explanatory arguments for the decision reached by a machine learning program. The problem with this approach is that it will provide arguments to justify the 20% or so of incorrect decisions.

5.2 Evaluation of Arguments

5.2.1 Abstract Accounts of Argumentation

In the 1990's Dung's abstract account of argumentation was mainly used as the final stage of a three-stage model of argumentation: construction of arguments, identifying their conflict relations and resolving the conflicts with preference information [Prakken, 1995]. This results in a set of arguments with a defeat relation, to which any semantics of [Dung, 1995] can be applied. Around 2000 an alternative approach emerged, in which arguments are directly encoded in abstract argumentation frameworks [Bench-Capon, 2002b] and in which preference information that is needed to resolve conflicts is added to these abstract frameworks after they have been constructed [Amgoud and Cayrol, 2002, Bench-Capon, 2003]. Briefly, the idea is to say that if argument A attacks argument B and is not inferior to B, then A defeats B. The semantics of abstract argumentation frameworks is then applied with this defeat relation. This idea is explicit in [Amgoud and Cayrol, 2002] and is indirectly modelled by [Bench-Capon, 2003] by attaching a (legal, ethical or societal) value to each argument and evaluating the success of attacks in terms of an ordering on the set of values.

In a series of subsequent papers Atkinson, Bench-Capon and colleagues applied the latter approach to frameworks where the arguments instantiate practical-reasoning argument schemes and in which the critical questions of these schemes are pointers to attacking arguments [Atkinson et al., 2005, Atkinson and Bench-Capon, 2005, 2007]. This approach is very attractive as long as arguments do not have an internal inferential structure, because of the simplicity and elegance of the theory of abstract argumentation frameworks. However, [Modgil and Prakken, 2013] argue that when arguments do have an internal inferential structure, an explicit account of the structure of arguments should be given, in order to apply the preference information to the points at which the arguments conflict. Thus the use of preference information to resolve conflicts comes before the generation of abstract argumentation frameworks. This approach is formalised in the ASPIC+ framework.

5.2.2 Values

[Bench-Capon, 2002b]'s addition of values to abstract argumentation frameworks was inspired by a preceding body of work of himself and others on the use of values in models of case-based reasoning [Bench-Capon and Sartor, 2001], [Bench-Capon, 2002a] [Prakken, 2002] [Sartor, 2002], which work was in turn inspired by [Berman and Hafner, 1993]. Criticising purely factor-based models of case-based reasoning, Berman and Hafner argued that often a factor can be said to favour a decision by virtue of the purposes served or values promoted by taking that decision because of the factor. A choice in case of conflicting factors is then explained in terms of a preference ordering on the purposes, or values, promoted or demoted by the decisions suggested by the factors[18]. Cases can then be compared in terms of the values at stake rather than on the factors they contain.

The role of purpose and value is often illustrated with some well-known cases from Anglo-American property law on ownership of wild animals that are being chased. Here we follow the analysis of three of these cases given by [Bench-Capon, 2002a]. In *Pierson* plaintiff was hunting foxes for sport on open land when defendant killed the chased fox and carried it away. The court held for defendant. In *Keeble* a pond owner placed a duck decoy in his pond with the intention to sell the caught ducks for a living. Defendant used a gun to scare away the ducks, for no other reason than to damage plaintiff's business. Here the court held for plaintiff. Finally, in *Young* both plaintiff and defendant were fishermen fishing in the open sea. Just before plaintiff closed his net, defendant came in and caught the fishes with his own net.

Let us assume that the task is to argue for a decision in *Young* on the basis of *Pierson* and *Keeble*. If cases are only compared on the factors they contain, then no ruling precedent can be found. *Young* contains pro-plaintiff factors absent in *Pierson*, namely, that the plaintiff was pursuing his livelihood, so *Young* can be distinguished from *Pierson*. Moreover, *Young* lacks a pro-plaintiff factor of *Keeble*, namely, that the plaintiff was hunting on his own land, and contains a pro-defendant factor that is not in *Keeble*, namely, that the defendant was also hunting for a living. So *Young* can also be distinguished from *Keeble*.

However, Berman & Hafner convincingly argue that skilled lawyers do not confine themselves to factor-based comparisons, but often frame their arguments in terms of the values that are at stake.[19] [Bench-Capon, 2002a] applies this view to the above cases and assumes that three values are at stake in these cases, viz. economic benefit for society

[18]The need for values to resolve issues requiring choice is also found in [Perelman et al., 1980] and Searle [Searle, 2003].

[19]Below we will use 'values' to cover also purposes, policies, interests etc.

(*Eval*), legal certainty (*Cval*), and the protection of property (*Pval*). Then a key idea is to specify how case decisions advance values.

- Deciding for a side because that side was hunting for a living advances *Eval*.
- Deciding for a side because that side was hunting on his own land advances *Pval*.
- Deciding for a side because that side had caught the animal advances *Pval*.
- Deciding for a side because the other side had not caught the animal advances *Cval*.

We can then say that *Pierson* was decided for defendant to promote legal certainty and since no values are served by deciding for plaintiff: he was not hunting for a living so economic benefit would not be advanced, and he had not yet caught the fox and was hunting on open land, so there are no property rights to be protected. Further, we can say that *Keeble* was decided for plaintiff since the value of economic benefit and the protection of property are together more important than the value of certainty. Thus *Keeble* also reveals part of an ordering of the values. Finally, in this interpretation of *Pierson* and *Keeble*, *Young* should be decided for defendant: the value of economic benefit does not support plaintiff since defendant was also fishing for his living, the value of protecting property does not apply since plaintiff had not yet caught the fish and was not on his own land, so the only value at stake is certainty, which is served by finding for the defendant.

We now give a general argument-scheme account of the reasoning involved, which captures the essence of how the above-cited papers analyse these cases and which can be formalised in ASPIC+ along the lines of [Bench-Capon et al., 2013]. The first idea is that the specification of how case decisions advance values can be used in the following argument scheme.

Argument scheme from case decisions promoting values

Deciding *Current Pro* promotes set of values V_1
Deciding *Current Con* promotes set of values V_2
V_1 is preferred over V_2

Therefore (presumably), *Current* should be decided *Pro*.

Here *Pro* and *Con* are variables ranging over {*Plaintiff*,*Defendant*}. Another idea is that whether a set of values is preferred over another set of values, can be derived from a precedent (as in our example from *Keeble*).

Argument scheme from preference from precedent

Deciding *Precedent Pro* promotes set of values V_1
Deciding *Precedent Con* promotes set of values V_2
Precedent was decided *Pro*

Therefore (presumably), V_1^+ is preferred over V_2^-

Here the notation V_1^+ denotes any superset of V_1 of values while V_2^- denotes any subset of V_2. This notation captures *a fortiori* reasoning in that if in a new case deciding *Pro* promotes at least V_1 and possibly more values, while deciding *Con* promotes at most V_2, then the new case is even stronger for *Pro* than the precedent.

If it is also given that a proper superset of values is always preferred over a proper subset, then the first scheme directly applies to *Young*, since deciding *Young* for the defendant promotes $\{Pval, Eval\}$ while deciding *Young* for the plaintiff promotes $\{Eval\}$. However, imagine another new case in which deciding for the plaintiff promotes $\{Pval, Eval\}$ or a superset thereof, while deciding for the defendant promotes $\{Cval\}$: then the second scheme is needed to infer the preference of the first value set over the second (for instance, from *Keeble*), after which the first scheme can be applied to conclude that the plaintiff should win.

[Bench-Capon and Sartor, 2001, 2003] employ a similar way to express that factor-decision rules promote values, and a similar way to derive rule preferences from the preference ordering on the sets of values they promote. But then they embed this in a method for constructing theories that explain a given set of cases, inspired by [McCarty, 1976]'s view of legal case-based reasoning as theory construction (see Section 4.1.6 above). Theory construction is modelled by Bench-Capon and Sartor as an adversarial process, where both sides take turns to modify the theory so that it explains the current case in the way they want. The process starts with a set of factor-value pairs and a set of cases represented in terms of factors and an outcome. Then the theory is constructed by creating rules plus rule priorities derived from value preferences. This continues until the theory can be applied to give an outcome for the case under consideration. At this point the onus moves to the other party, who must attempt to extend the theory to produce a better theory with an outcome for its favoured side, whereupon, it is again the turn of the original side. This process of extending and refining the theory continues until there is no possible extension of the theory which changes the outcome.

This approach was tested empirically in [Chorley and Bench-Capon, 2005b] and [Chorley and Bench-Capon, 2005a]. The first of these papers explored the use of CATE (CAse Theory Editor) in a series of experiments intended to explore a number of issues relating to the theories constructed using the operators of [Bench-Capon and Sartor, 2003], including how the theories should be constructed, how sets of values should be compared, and the representation of cases using structured values (which are akin to dimensions) as opposed to factors. In CATE, the construction of theories is done by the user, supported by the CATE toolset.

The second paper described AGATHA (Argument Agent for Theory Automation) which was designed to automate the theory construction process, by constructing the theory first as a search over the space of possible theories, and then as a two player dialogue game (which could be played with the AGATHA program playing both sides). A set of search operators and argument moves are defined in terms of the theory constructors and the resulting theories are evaluated according to their explanatory power and their simplicity. The search or game continues until it is not possible to produce a better theory. Several search methods were investigated: brute force and heuristic search using A* and adversarial search using α/β pruning. The results proved to be good, as reported in [Chorley and Bench-Capon, 2005a]:

> AGATHA produces better theories than hand constructed theories as reported in [Chorley and Bench-Capon, 2005b], and theories comparable in explanatory power to the best performing reported technique, IBP [Ashley and Brüninghaus, 2009]. Note also that AGATHA can be used even when there is no accepted structural model of the domain, whereas IBP relies on using the structure provided by the Restatement of Torts.

The attention for the role of value and purpose led to accounts of legal interpretation as a decision problem, namely, as a choice between alternative interpretations on the basis of the likely consequences of these interpretations in terms of promoting and demoting values.

Bench-Capon and Atkinson have studied legal practical reasoning in the context of [Bench-Capon, 2003]'s value-based abstract argumentation frameworks. As explained above in Section 5.2.1, such VAFs extend abstract argumentation frameworks by giving each argument a value that it promotes and by defining a total ordering on these values. Attacks are then resolved by comparing the relative preference of the values promoted by the conflicting arguments. In e.g. [Atkinson et al., 2005, Atkinson and Bench-Capon, 2005] the instantiation is studied of the arguments in VAFs with the following so-called argument scheme for practical reasoning:

In the current circumstances R
Action A should be performed
To bring about new circumstances S
Which will realise goal G
And promote value V

The scheme comes with a list of critical questions that can be used to critique each element of this scheme and to generate counterarguments to uses of the scheme. For example:

CQ1:	Are the believed circumstances true?
CQ2:	Assuming the circumstances, does the action have the stated consequences?
CQ7:	Are there alternative ways of promoting the same value?
CQ9:	Does doing the action have a side effect which demotes some other value?
CQ16:	Is the value indeed a legitimate value?

This generates a VAF as follows. Instantiations of this scheme are arguments, while arguments for incompatible actions and 'bad' answers to critical questions are counterarguments to such arguments. Then each argument is assigned a value.

Atkinson and Bench-Capon have applied this approach both to legal interpretation and to legislative and policy debates. In [Atkinson

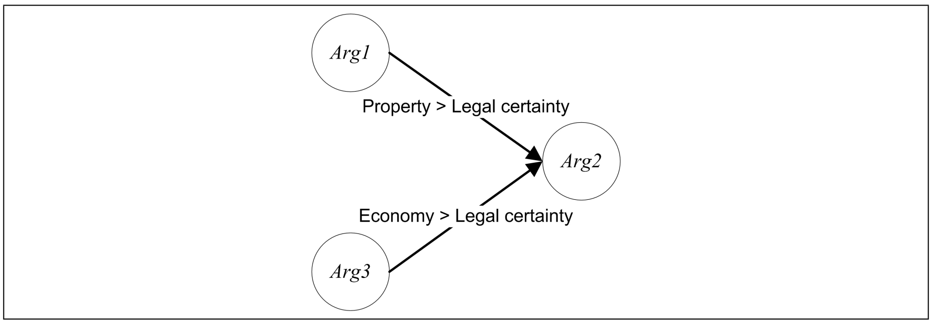

Figure 9: Dung-style AF for the wild animals case

et al., 2005] they applied it to reasoning with precedents, representing the *Keeble* case as follows (note that they equate circumstances S and goal G):

> **Arg1**:
> Where plaintiff is hunting on his own land
> find ownership established
> as plaintiff's property is thus respected
> which promotes the protection of property
>
> **Arg2**:
> Where plaintiff is hunting for a living
> find ownership established
> as plaintiff's activities are thus encouraged
> Which promotes the economy
>
> **Arg3**:
> Where there is no possession
> find ownership not established
> as this will reduce litigation
> which promotes legal certainty.

Here both Arg1 and Arg3 and Arg2 and Arg3 attack each other. To explain the decision in *Keeble*, the values of protection of property and the economy should be preferred to the value of legal certainty, so that Arg3 is defeated by either Arg1 and Arg2. The resulting abstract argumentation framework is displayed in Figure 9.

While this approach has its merits, it also has some limitations. First, it does not deal naturally with aggregation of values promoted by the same decision, unlike the above-discussed *Argument scheme from case decisions promoting values.* Second, different parts of the scheme model different kinds of inference steps. That action A will result in consequences S is (causal) epistemic reasoning, while the step to the value is evaluative and the conclusion that A should be performed is practical reasoning. Now a conflict on whether the action has a certain result is different from a conflict on whether the action should be performed. The latter indeed requires value comparisons but the former is a conflict of epistemic reasoning, to which value considerations do not apply.

An alternative approach is to formulate practical reasoning as a combination of various elementary argument schemes and to embed their use in a framework for argumentation that allows for the stepwise construction of arguments. This approach, briefly discussed above, was applied in the next decade, to be discussed in Section 6 below.

5.2.3 Burden of Proof

Above we saw that [Freeman and Farley, 1996] and [Gordon et al., 2007, Gordon and Walton, 2009] incorporate standards of proof in their models of legal argument. However, [Prakken and Sartor, 2009] argue that standards of proof cannot be applied on their own but are relative to *burdens* of proof, and different phases of a legal proceeding can be about different proof burdens. Generally[20] a distinction is made between a burden to provide evidence on an issue during a proceeding (in common-law systems often called the burden of production) and a burden to prove that a claim is true or justified beyond a given standard of proof (in common-law systems often called the burden of persuasion). If the burden of production on an issue is not met, the issue is decided as a matter of law against the burdened party, while if it is met, the issue is decided in the final stage of the proceeding according to the burden of persuasion. In the law the burdens of production and persuasion are usually determined by the 'operative facts' for a legal claim, i.e. the facts that legally are ordinarily sufficient reasons for the claim. The

[20]Much of this Section is taken from [Prakken and Sartor, 2015a].

law often designates the operative facts with rule-exception structures. For instance, for manslaughter the operative facts are that there was a killing and that it was done with intent, while an exception is that it was done in self-defence. Therefore, at the start of a criminal proceeding, the prosecution has the burden to produce evidence on 'killing' and 'intent'; if this burden is fulfilled, the defence's burden to produce evidence for 'self-defence' is activated. For operative facts the burdens of production and persuasion usually go together so in our example the prosecution also has the burden of persuasion for 'killing' and 'intent'. However, for exceptions things are more complicated. In criminal proceedings usually the defence only has a burden of production for an exception while if fulfilled, the prosecution then has an active burden of persuasion against the exception. For instance, once the defence has produced evidence for 'self-defence', the prosecution has the burden of persuasion that there was no self-defence. By contrast, in civil cases often the burden of persuasion holds for an exception also: for instance, in Dutch and Italian law insanity at the time of accepting an offer is an exception to the rule that offer and acceptance create a binding contract, but if the evidence on insanity is balanced, the party claiming insanity will lose on that issue.

This account fits rather well with argumentation-based logics for defeasible reasoning. The idea is that a burden of persuasion for a claim is fulfilled if at the end of a proceeding the claim is sceptically acceptable according to the argumentation logic applied to the then available evidence [Prakken and Sartor, 2009]. However, there is a complication, namely, the just-mentioned possibility in civil cases that the burden of persuasion is distributed over the adversaries. The complication can best be explained in terms of abstract argumentation frameworks. Consider again the above contract example, and consider the following arguments:

> P_1: The contract was concluded because there was an offer and acceptance (assuming there is no exception)
>
> O_1: There is an exception since the offeree was insane when accepting the offer (evidence provided)
>
> P_2: The offeree was not insane when accepting the offer, since (evidence provided)

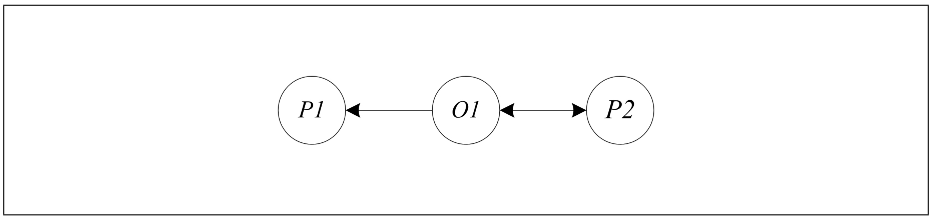

Figure 10: A Dung graph

It seems reasonable to say that argument O_1 strictly defeats P_1, since it refutes P_1's assumption that there is no exception. Assume, furthermore, that O_1 and P_2 are regarded as equally strong (according to any suitable notion of strength). Then it seems reasonable to say that both arguments defeat each other. The resulting Dung graph is displayed in Figure 10:

The grounded extension is empty, while two preferred extensions exist: one with P_1 and P_3 and one with O_1. So the plaintiff has no sceptically acceptable argument for his main claim. Yet according to the law the plaintiff wins, since the defendant has not fulfilled her burden of persuasion as regards her insanity: O_1 is also just defensible.

This is one challenge for a Dung-style approach. Another challenge is to account for the fact that different kinds of legal issues can have different standards of proof. For example, in common-law jurisdictions claims must in criminal cases be proven 'beyond reasonable doubt' while in civil cases usually proof 'on the balance of probabilities' suffices. Consider now again the killing-in-selfdefence example, and assume that the prosecutor has an argument P_1 that the accused killed, the accused has an argument O_1 that he killed in selfdefence, and the prosecutor has an argument P_2 against this argument, which is considerably stronger than its target but not strong enough to satisfy the 'beyond reasonable doubt' proof standard. In a Dungean account defeat is an all-or-nothing matter, so to obtain the legally correct outcome that the accused must be acquitted, in this case O_1 and P_2 must be said to defeat each other (resulting in Figure 10).

One approach to deal with these problems is to give up a Dungean approach. As we saw above in Section 4.2.4, this is what Tom Gordon did in his Carneades system. However, as noted above, his approach arguably conflates the distinction between proof standards and proof burdens. Prakken and Sartor have made various attempts to deal with these problems in a Dungean setting, e.g. in [Prakken and Sartor, 2011, Calegari et al., 2021]. Their most recent attempt is [Prakken and Sartor, 2023], in which they use ASPIC+ as a metalevel formalism to specify decompositions of reasoning problems, where each subproblem can be solved by its own reasoning method, which can be of any kind. Then shifts in the burden of persuasion can be modelled in metalevel rules that explicitly indicate which propositions should be proven, and degrees of defeat can be modelled in evidential problem-solving modules that apply some numerical model of reasoning under uncertainty, such as Bayesian probability theory.

A concept closely related to that of burden of proof is the notion of **presumption**. Legal presumptions obligate a fact finder to draw a particular inference from a proved fact. Typical examples are a presumption that the one who possesses an object in good faith is the owner of the object, or a presumption that when a pedestrian or cyclist is injured in a collision with a car, the accident was the driver's fault. Some presumptions are rebuttable while others are irrebuttable.

The logical interpretation of (rebuttable) presumptions is less complicated than for burdens of proof but not completely trivial. [Prakken and Sartor, 2008] argue that the function of legal presumptions is not to *allocate* a proof burden but to *fulfil* it. More precisely, the interpret presumptions are default rules or default conditionals, which can be used in arguments for claims that have a proof burden attached to them.

Further logical issues concerning presumptions and burdens of proof are discussed in [Prakken and Sartor, 2007a, 2008, 2009].

5.2.4 Accrual

One recurring theme in the computational study of argumentation is that of accrual of arguments, or how several arguments for the same conclusion should be combined. This issue has been especially (although not exclusively) been studied in the context of AI & Law. The main issue is whether accrual should be modelled at the knowledge representation level, by combining different reasons for the same conclusion in antecedents of rules (in [Prakken, 2005b] called the *KR approach*), or whether it should be modelled at the logical level as a logical operation on arguments (in [Prakken, 2005b] called the *inference approach*). Early work on the inference approach was [Verheij, 1996]'s Cumula system and [Hage, 1996]'s Reason-Based Logic. [Prakken, 2005b] proposed three principles that any model of argument accrual should satisfy, namely:

- An accrual is sometimes weaker than its elements (since reasons can interact, as in 'both heat and rain are a reason not to go jogging, but the combination is so pleasant that it is a reason to go jogging');
- an accrual makes its elements inapplicable (for instance, if it is hot and rainy, then the individual fact that it is hot cannot be used any more in an argument);
- flawed reasons or arguments may not accrue (for instance, if the argument that it rains can be refuted, then the argument 'it rains, therefore I should not go jogging' cannot be accrued with the argument 'it is hot, therefore I should not go jogging'; only the latter argument should be considered).

[Prakken, 2005b] then proposed an inference-based model that satisfied these three principles in terms of a combination of Dung's theory of abstract of argumentation frameworks with [Pollock, 1995]'s theory of defeasible reasons. The key idea was to label conclusions of individual arguments and to have a defeasible accrual reason $\varphi^{l1}, \ldots, \varphi^{ln} \Rightarrow \varphi$ that can be applied to the conclusions of a set of arguments with the same conclusion when unlabelled.

5.3 Use of Arguments

5.3.1 Dialogue Games

Research on dialogue systems continued in this decade. Partly motivated by the earlier AI and Law work on dialogue systems, [Prakken, 2005a] proposed a general framework for specifying systems for two-party persuasion dialogue, and then instantiated it with some example protocols. The framework largely abstracts from the logical language, the logic and

Acts	Attacks	Surrenders
claim φ	*why* φ	*concede* φ
φ *since* S	*why* $\psi (\psi \in S)$ φ' *since* S' (φ' *since* S' defeats φ *since* S)	*concede* ψ $(\psi \in S)$ *concede* φ
why φ	φ *since* S	*retract* φ
concede φ		
retract φ		

Table 1: An example communication language [Prakken, 2005a]

the communication language but the logic is assumed to be argument-based (in fact a preliminary version of the ASPIC+ framework) and to conform to grounded semantics.

A main motivation of the framework is to ensure focus of dialogues while yet allowing for freedom to move alternative replies and to postpone replies. This is achieved with two main features of the framework. Firstly, an explicit reply structure on the communication language is assumed, where each move either *attacks* or *surrenders to* its target. An example language of this format is displayed in Table 1. Secondly, winning is defined for each dialogue, whether terminated or not, and it is defined in terms of a notion of *dialogical status* of moves. The *dialogical status* of a move is recursively defined as follows, exploiting the tree structure of dialogues. A move is *in* if it is surrendered or else if all its attacking replies are *out*. (This implies that a move without replies is *in*). And a move is *out* if it has a reply that is *in*. Then a dialogue is (currently) won by the proponent if its initial move is *in* while it is (currently) won by the opponent otherwise.

Together, these two features of the framework allow for a notion of relevance that ensures focus while yet leaving the desired degree of freedom: a move is *relevant* just in case making its target *out* would make the speaker the current winner. Termination is defined as the situation that a player is to move but has no legal moves.

[Prakken et al., 2005] applied the framework to specify a protocol for dialogues about who has the burden of proof for a given claim. [Prakken, 2008a] extended an instance of the framework with a neutral third party in order to model so-called adjudicator dialogues, in which an adjudicator monitors whether the adversaries respect the protocol and in the end decides the dispute. The main feature of the model is a division into an argumentation phase, where the adversaries plea their case and the adjudicator has a largely mediating role, and a decision phase, where the adjudicator decides the dispute on the basis of the claims, arguments and evidence put forward in the argumentation phase. The model allows for explicit decisions on admissibility of evidence and burden of proof by the adjudicator in the argumentation phase. Adjudication is modelled as putting forward arguments, in particular undercutting and priority arguments, in the decision phase. [Prakken, 2008b] applied this model in a case study to a Dutch civil ownerships dispute.

When a dialogue protocol is fully specified in some formal language, then its metatheory can be investigated with the help of automated reasoning tools. [Brewka, 2001] specified his protocols in a dialect of the situation calculus and [Artikis et al., 2003] formalised variations of Brewka's protocols in the C^+ language of [Giunchiglia et al., 2004]. They then used implemented tools to verify various properties, such as the minimal length of dialogues that reach a certain state given a certain initial state. Another benefit of a logical formalisation of a dialogue protocol is that this supports the automatic execution of protocols. To this end, [Bodenstaff et al., 2006] formalised an instance of [Prakken, 2005a]'s framework in [Shanahan, 1999]'s version of the 'full' Event Calculus and then implemented it as a Prolog program. The implementation computes in any state of a dialogue the players' commitments, whether the moves made were legal, who is to move and what are the legal next moves. It can thus be used as a 'dialogue consultant' by a player, adjudicator or external observer.

Another strand of work was the dialogue protocols developed at the University of Liverpool. [Atkinson et al., 2006a] embedded [Atkinson et al., 2005, Atkinson and Bench-Capon, 2005]'s modelling of practical-reasoning argument schemes in a dialogue protocol in which the critical questions of the schemes drive the dialogue.

[Wardeh et al., 2009] used datamining for extraction association rules from case bases concerning the classification of routine claims for a hypothetical welfare benefit. They defined a dialogue game for refining the mined association rules through a dialogue with moves based on case-based reasoning systems such as CATO (see Section 4.1.5 above), including moves for citing, distinguishing, giving counter examples, and for pointing out unwanted consequences of a rule. The main idea is that during the course of the dialogue the rule is refined so that when the dialogue was complete, the winning rule is available to justify and explain an outcome. The authors also defined game strategies and tactics.

5.3.2 Dynamics of Case Law

It is a feature of case law that it evolves over time, with decisions being refined, and overruled in response to novel fact situations and social practices and values. It had received some attention in AI and Law (e.g. [Rissland and Friedman, 1995] and [Berman and Hafner, 1995]). In [Henderson and Bench-Capon, 2001] argumentation was used to address the topic. The paper was based on some remarks by Levi [Levi, 1948]:

> "Reasoning by example shows the decisive role which the common ideas of the society and the distinctions made by experts can have in shaping the law. The movement of common or expert concepts into the law may be followed. The concept is suggested in arguing difference or similarity in a brief ... In subsequent cases, the idea is given further definition and is tied to other ideas which have been accepted by courts. It is no longer the idea which was commonly held by society"

The particular domain was the notion of whether a person owes a duty of care in virtue of their occupation. The idea was to start from a "common sense" ontology of occupations. Then given a set of cases stating whether occupation owed a duty of care or not arguments could be constructed for new occupations. Arguments for could be found by finding the closest common ancestors covering both the current case and a pro-case and a contra-case. Arguments for and against could then be generated by pointing to the similarities and differences between the current and previous cases. The decision indicates which arguments are accepted and hence which rules are established, thus developing a specifically legal concept. As more cases are decided these rules are narrowed and broadened and exceptions identified. The implemented program illustrated this process of reinterpretation and modification. One result was that the order in which cases were presented was shown to matter: given a different sequence, different arguments will be available, and the final rule may be different also.

5.3.3 Evidence

Work in AI and Law on evidential reasoning started in 2003 with articles by [Keppens and Zeleznikow, 2003], [Bex et al., 2003] and [Prakken, 2004]. The latter two of these focus on (logical) evidential argumentation, where the knowledge base contains the evidence in a case and non-legal, common-sense rules are used to reason towards a conclusion. [Bex et al., 2003] discuss a number of argumentation schemes for evidential reasoning, such as the *scheme from appeal to expert opinion* (Section 5.1.1), the *scheme from appeal to general knowledge* and the *scheme from appeal to witness testimony*. Furthermore, they also discuss general schemes for inferences from perception or memory ([Pollock, 1995]), and their relation to, for example, the witness testimony scheme. Inspired by [Loui and Norman, 1995] (Section 4.1.2), [Bex and Prakken, 2004] show that the witness testimony scheme "if a witness testifies that P is the case then usually P is the case" can be unpacked into "if a witness testifies that they observed P then usually they believe that they observed P", "if a witness believes that they observed P then usually their senses gave evidence of P" and "if a witness' senses gave evidence of P then usually P is the case".

Where [Bex et al., 2003] focus mostly on arguments from evidence towards some conclusion, [Keppens and Zeleznikow, 2003] use scenarios (or stories, cf. Section 5.1.2) to explain the evidence. [Bex et al., 2006, 2007a, 2010] combined argument-based and story-based reasoning in one hybrid theory, where arguments based on evidence can be used to support or attack explanatory stories. Figure 11 shows how the two stories explaining Mary's death can be supported and attacked by arguments based on evidence. For story S1, we have two supporting arguments based on a witness testimony and expert DNA evidence. Story S2 is

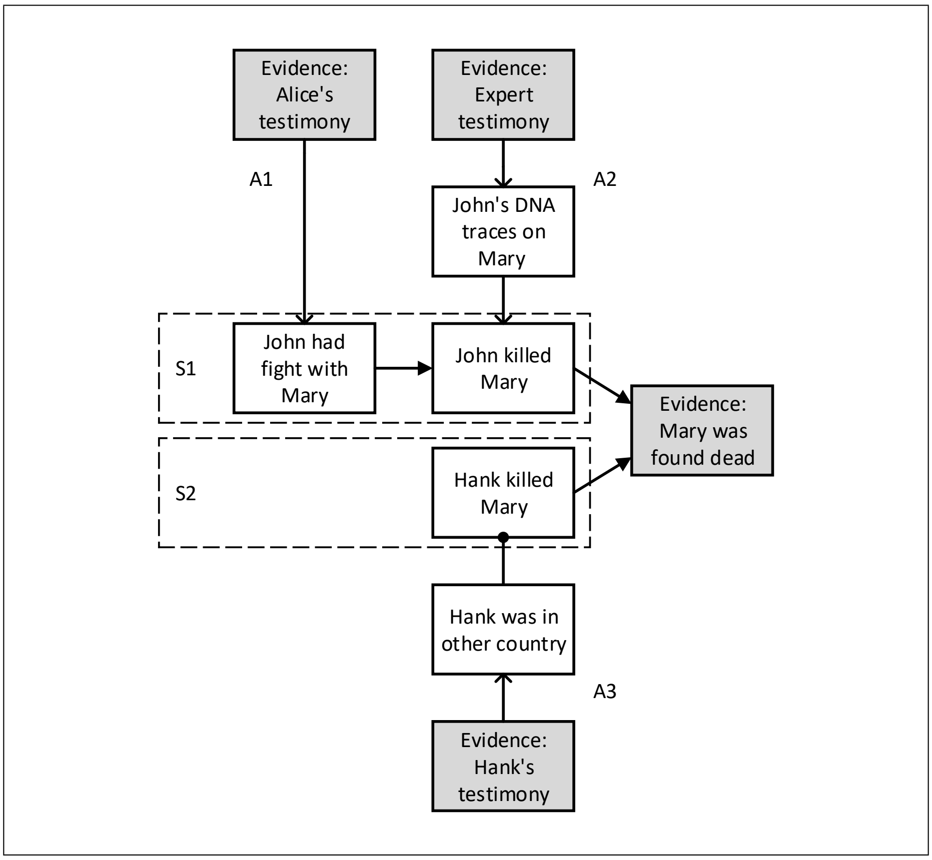

Figure 11: Arguments based on evidence supporting (A1, A2) and attacking (A3) the two stories about Mary's death.

attacked by the fact that Hank was in another country as testified by himself. The work on the hybrid theory was used as a basis for sense-making using argument and scenario diagramming in the AVERs tool [van den Braak et al., 2007, Bex et al., 2007b] (see also Section 5.3.4).

5.3.4 Implemented argument structure and evaluation

Following the work in the 1990s on argumentation and dialogue software (Section 4.3.4), work continued on systems for various argumentation tasks, often using formalized graphical representation formats.

Verheij [2003a] continued the work on the ArguMed system (see Section 4.3.4) by extending its expressiveness and developing a corresponding formalization of the logic of argumentation (DefLog, Verheij [2003b]). In the resulting system (ArguMed based on DefLog), the graphical elements in a diagram correspond to formal sentences: each box (representing a statement) corresponds to an elementary proposition in the logic, and each arrow to a conditional sentence. There are conditional representations of supporting and attacking reasons, and the conditional relations can themselves be supported and attacked (using nested conditional sentences). By this mechanism, undercutting defeaters attacking the connection between a reason and the conclusion it supports [Pollock, 1995] can be modelled as an attack on a conditional and warrants supporting the connection between a reason and its conclusion [Toulmin, 1958] as support of a conditional. The software computes which statements are justified and which defeated by evaluating the *prima facie* assumptions in the system (using a formal generalization of the stable semantics of abstract argumentation frameworks [Dung, 1995]). For instance, the example argument in the introduction (Figure 1) can be represented in ArguMed using the following formal sentences:

`Mary is original owner` (an elementary sentence)
`John is the buyer`
`if(Mary is original owner, Mary is owner)`
(a conditional sentence expressing support)
`if(John is the buyer, John is owner)`
`if(John was not bona fide, x(if(John is the buyer, John is owner)))`
(a conditional sentence expressing attack, indicated using the so-called dialectical negation `x`)
`inc(John is owner, Mary is owner)`
(abbreviating incompatibility, i.e. that the two staments attack each other)
`if(John bought the bike for 20 euros, John was not bona fide)`
`John bought the bike for 20 euros`
`if(John bought the bike for 25 euros, x(John bought the bike for 20 euros))`

```
John bought the bike for 25 euros
if(John bought the bike for 25 euros, John was
                                  not bona fide)
```

Evaluating this set of sentences as *prima facie* assumptions in DefLog, gives the result visualized in Figure 1: all sentences listed above are justified, except for the sentences `John bought the bike for 20 euros` and `if(John was not bona fide, x(if(John is the buyer, John is owner)))`, which are defeated sentences in the *prima facie* theory. The sentences `Mary is owner` and `John was not bona fide` are justified by derivation from other justfied sentences (using Modus Ponens) and sentences `John is owner` and `John bought the bike for 20 euros` are defeated by derivation from other justfied sentences.

The Carneades model of argument [Gordon et al., 2007] continued on the formalization of argument structure and evaluation. The model design was aimed as the basis for software development. A characteristic feature of the model is that it included proof standards. For each premise, a burden of proof can be allocated to a proponent and opponent. See Section 5.1.4 for an extended discussion.

Ashley et al. [2007] study the use of argument diagramming in the context of intelligent tutoring systems. The system described, LARGO, is developed to train the legal reasoning skills of first year law students. The focus is on hypothetical legal reasoning on the basis of US Supreme Court cases. Students can propose a rule deciding a case, challenge such a rule, and continue the discussion by proposing analogies or distinctions (cf. also HYPO, see Section 3.1.1). The system includes domain-specific critical questions thereby allowing for system feedback on weaknesses in an argument move. The system's user interface provides a graphical representation of the argument structure.

The AVERs system [van den Braak et al., 2007] was developed in order to support the process of making sense of the evidence collected in crime investigation. The approach is hybrid in the sense that it combines argumentation and scenario elements. Initially observed facts can be connected to hypothesized events, and combined into stories. The elements of stories can be supported by evidence using arguments. The AVERs system supports the visualization of hybrid diagrams of events and evidence connected by arguments. The system is connected to the work on a hybrid combination of arguments, stories and evidence discussed in Section 5.3.3 [Bex et al., 2007a, Bex, 2011].

Verheij [2007] discusses boxes-and-arrows diagramming in the context of argumentation support software, discussing both opportunities and limitations. Suggestions to go beyond boxes-and-arrows diagram include the need for more expressiveness than what boxes and arrows allow, the inclusion of argumentation schemes (see Section 5.1.1), refocusing on natural language, and simplified diagram structures that may be more helpful.

5.3.5 Web Based Dialogues

Enabling citizens to engage in dialogue with their governments is an important feature of a democracy. Whilst this had for some time been conducted by traditional means, such as writing letters, attending town hall debates and holding individual local ‘surgeries’, new web-based methods of interaction have been developed to exploit emerging digital technologies. A number of such tools have been developed that make use of computational models of argument. One key example is the Parmendies tool [Atkinson et al., 2006b], where the aim was to present to members of the public a policy proposal for their review and critique.

The policy was presented as an instantiation of the practical reasoning argumentation scheme with values [Atkinson et al., 2006a], discussed earlier in this section. Using the scheme enabled presentation of the current situation the policy scenario was arising in and what the policy proposed was meant to achieve in terms of facts, goals and values. This policy proposal was all presented to the user through a webpage. They were then given the opportunity to critique the policy in terms of relevant critical questions characteristic of the practical reasoning argumentation scheme. The critical questions were posed systematically through navigation to subsequent webpages, to tease apart the precise points of disagreements and motives for these that a user may wish to express about the policy. Different people might disagree with how the current policy situation was expressed, others might question whether the policy would achieve the intended ends, and yet others might oppose these ends because the do not subscribe to the values the ends promote.

The tool was thus intended to enable a form of web-based dialogue between citizens and policy makers. Parmenides later formed the basis for the development of a richer 'Structured Consultation Tool (SCT)' [Bench-Capon et al., 2015] produced as part of the IMPACT project[21].

5.3.6 Game Theory

In a dialogical setting issues of strategy and choice naturally arise but in a legal context they have not been much investigated. In [Riveret et al., 2007, Roth et al., 2007, Riveret et al., 2008] game theory was applied to the problem of determining optimal strategies in adjudication debates (see also [Sartor et al., 2009]). In such debates, a neutral third party (for example, a judge or a jury) decides at the end of the debate whether to accept the statements that the opposing parties have made during the debate, so the opposing parties must make estimates about how likely it is that the premises of their arguments will be accepted by the adjudicator. Moreover, they may have preferences over the outcome of a debate, so that optimal strategies are determined by two factors: the probability of acceptance of their arguments' premises by the adjudicator and the costs/benefits of such arguments. In [Riveret et al., 2007] the logical basis is Defeasible Logic [Antoniou et al., 2000]; in [Roth et al., 2007] it is a dynamic version of the argument game of [Prakken and Sartor, 1997]; and in [Riveret et al., 2008] an abstract argument game.

6 2010s: Computational Argumentation as a Field

By the 2010s, argumentation approaches have become influential in Artificial Intelligence and attempts are made at standardization. This decade saw a deepening of understanding of many themes and topics, such as machine learning and evidence. In the study of the latter, Bayesian networks and other probabilistic approaches receive more attention.

[21]Integrated Method for Policy Making Using Argument Modelling and Computer Assisted Text Analysis, in the European Framework 7 project (Grant Agreement No 247228) in the ICT for Governance and Policy Modeling theme (ICT-2009.7.3).

6.1 Argument Generation

6.1.1 Rule-Based Approaches

Governatori and colleagues continued their work on applying Defeasible Logic to legal reasoning [Governatori and Sartor, 2010, Governatori et al., 2013, Rotolo et al., 2015]. Satoh and colleagues developed an alternative rule-based approach close to logic programming called PROLEG [Satoh et al., 2012].

6.1.2 Argumentation Schemes

In the 2010s the idea of using argument schemes for case-based reasoning in the context of a general structured account of argumentation (see Section 5.2.2) was further developed. This was first done by [Bench-Capon and Prakken, 2010], who semiformally sketched how a collection of argument schemes involving value-based reasoning can be formalised in argumentation logics. They applied these schemes in a semiformal reconstruction of various US Supreme Court cases concerning the automobile exception to the constitutional protection against unreasonable searches and seizures. Next, [Prakken et al., 2015] (first online in 2013) gave a full formalisation in ASPIC+ of a set of argument schemes modelling CATO-style case-based reasoning. Building on this work, [Bench-Capon et al., 2013] added schemes for value-based reasoning, while [Atkinson et al., 2013] added schemes for reasoning with dimensions. In all this work the idea is that argument schemes can be formalised as defeasible rules in ASPIC+ (or a similar system) while critical questions are pointers to rebutting or undercutting counterarguments.

6.1.3 Machine Learning

Even though the focus in AI & Law was very much still on formal logical models, at least when it concerned argumentation in AI & Law, there were already a few authors that included machine learning in their work on argumentation.

[Ashley and Walker, 2013] use logical models of argument to represent legal rules and the reasoning from evidence to some (legal) conclusion. Thus, more high-level legal concepts are decomposed into facts (entities, events) that can be more readily mined from texts using machine learning methods. They further annotate such facts and confidence levels in a legal corpus about vaccine injury tort cases. Their ultimate aim is to develop a QA-system.

[Schraagen and Bex, 2018] also propose a QA-system based on argumentation to assist the Dutch Police in the assessment of crime reports submitted by civilians. Similar to [Ashley and Walker, 2013], they use a logical argumentation model to decompose legal concepts into facts, the latter of which can be gathered by extracting them from the initial user input using machine learning NLP ([Schraagen et al., 2017, Schraagen and Bex, 2019] provide explorations in this regard), or by asking the user relevant questions based on the conclusions that can be drawn from the argumentation model at any time. The aim of [Schraagen and Bex, 2018] was to learn policies for such question-asking using reinforcement learning. Further information about the ideas expressed in this paper are discussed in Section 6.1.3.

6.1.4 Case Models and Argument Validity

Verheij [2017a] develops the formal connections between arguments, rules and cases, building on the idea that legal argument has two typical kinds of backing, namely cases and rules. Cases are used as the formal semantics of rule-based arguments, in the sense that cases formally determine which rules and arguments hold ('are valid').

An argument from certain premises to certain conclusions can have one of three types of validity, given a so-called case model. An argument is coherent if there exists a case (in the given case model) in which the argument's premises and the conclusions both hold. A coherent argument is conclusive if in each case in which the premises hold also the conclusions hold. In a case model, cases also come with an ordering, for instance representing their exceptionality. A case lower in the ordering is more exceptional. Using the ordering relation, a third kind of validity can be defined, corresponding to the idea of defeasible reasoning: A coherent argument is presumptively valid if there is a case in which both the premises and conclusions hold, such that the case is at least as high in the ordering as each case in which the argument's premises hold.

The paper discusses formal analogy and distinction between cases

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	¬dut	¬dut	¬dut	dut	dut	dut	dut	dut	dut	dut	dut	dut	¬dut	¬dut	¬dut
¬dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg
	¬unl	unl	unl	unl	unl	unl	unl	unl	unl	unl	unl	unl	¬unl	¬unl	unl
		¬imp	imp	imp	imp	imp	imp	imp	imp	imp	imp	imp			imp
			¬cau	cau	cau	cau	cau	cau	cau	cau	cau	cau			cau
	¬vrt			vrt	vrt	vrt	¬vrt	¬vrt	¬vrt	¬vrt	¬vrt	¬vrt	vrt	¬vrt	
	¬vst			¬vst	¬vst	¬vst	vst	vst	vst	¬vst	¬vst	¬vst	¬vst	vst	vst
	¬vun			¬vun	¬vun	¬vun	¬vun	¬vun	¬vun	vun	vun	vun			
		¬ift		ift	¬ift	¬ift	ift	¬ift	¬ift	ift	¬ift	¬ift			
		¬ila		¬ila	ila	¬ila	¬ila	ila	¬ila	¬ila	ila	¬ila			
		¬ico		¬ico	¬ico	ico	¬ico	¬ico	ico	¬ico	¬ico	ico			
				¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	jus	jus	
				prp	prp	prp									¬prp

1 > 2 > 3 > 4 > 5 ~ 6 ~ 7 ~ 8 ~ 9 ~ 10 ~ 11 ~ 12 ~ 13 > 14 ~ 15 ~ 16

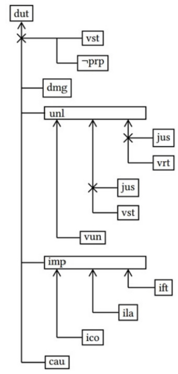

Figure 12: A case model (left) validating key arguments in Dutch tort law (right) [Verheij, 2017a]

(but not using the more fine-grained apparatus of factors and dimensions, as in HYPO; cf. Section 3.1.1). The approach also allows for the formalization of rebutting, undercutting and undermining attack.

A case model is presented that is the formal basis of a concrete, realistic legal domain, namely Dutch tort law about the duty to pay for damages (Figure 12). In that model, the least exceptional case (represented in the figure in the far left column, numbered 1) is a case in which there are no damages (¬`dmg`), so the question about a duty to pay damages does not arise. As a result, a presumptively valid conclusion is that there are no damages. Cases in which there are damages are more exceptional (the cases numbered 2 to 16 in the figure). The least exceptional cases in which there is a duty to pay for the damages (`dut`) correspond to the various combinations of conditions that determine such a duty (cases 5–13). An argument with such conditions as premises (e.g. `dmg` ∧ `unl` ∧ `imp` ∧ `cau` representing that there are damages by an unlawful act imputable to the actor causing the damages) presumptively justifies the conclusion that damages have to be paid. Cases with special circumstances, such as grounds of justification (`jus`, in cases 14 and 15), are more exceptional and do not imply a duty to pay. If such grounds are among the premises, there are so-called defeating circumstances attacking the argument to pay.

The case model approach suggests new ways to the formal integration of cases, rules and arguments, in which formalized cases provide a formal semantics for rule-based argument structure. Qualitative and quantitative representation results are given in [Verheij, 2016b]. Initial ideas of the case model approach to argument validity were developed in the context of evidence [Verheij, 2014, 2017b], further discussed in Section 6.4.3. The case model approach was also applied to the modeling of value-based argumentation [Verheij et al., 2016], focusing on time dependence as in the work by Berman and Hafner on New York tort cases (see Section 4.2.5). See Section 6.2.1 for other work on value-based argumentation in the 2010s.

6.2 Evaluation of Arguments

6.2.1 Values

Values received attention in a variety of different works by different groups of authors during this decade.

A series of papers by Grabmair and colleagues [Grabmair and Ashley, 2010] [Grabmair and Ashley, 2011], plus Grabmair's PhD thesis [Grabmair, 2016], feature values in a formal account of legal reasoning. The formal models set out in this suite of work capture the notion of one factual situation being preferable over another by virtue of the situations' respective effects on values. In [Grabmair and Ashley, 2010] legal sources are modelled as sets of value judgments and legal methodologies as collections of argumentation schemes. This model is then put to use to enable hypothetical reasoning within decision making on legal cases. The work was then extended in [Grabmair and Ashley, 2011] to further enable fine-grained case comparison whereby intermediate legal concepts were captured to determine their impact on the applicable values. A full 'value judgment formalism' is set out in detail to capture the interaction between facts and values, yielding the ability to produce arguments comparing cases within the task of legal case-based reasoning. Grabmair's PhD thesis [Grabmair, 2016] adds to the formalism an experimental implementation of the value judgment formalism and demonstrates how this implementation is capable of arguing about, and predicting outcomes of, cases from the CATO trade secret misappropriation dataset.

In a separate strand of work on values, [Sartor, 2010] studied how legal choices, and in particular legislative determinations, need to consider multiple rights and values, and can be assessed accordingly. Recognising that legal norms often prescribe the pursuit of conflicting goals, Sartor sets out a model of teleological reasoning through which legislative action is guided not only by constitutional 'action-norms', but also by constitutional 'goal-norms', that inform the legislator's teleological reasoning about which values should be advanced. The formal model provided is intended to capture the space of legislative and administrative actions by evaluating the teleological appropriateness of legislative choices.

With the increasing prevalence of multi-agent systems being deployed in real world scenarios, the work in [Bench-Capon and Modgil, 2017] advocates for such agents being equipped with the ability to reason about a system's norms, achieved by reasoning about the social and moral values that norms are designed to serve. A specific focus is placed on reasoning in circumstances where it can be argued that the rules should be broken and a decision should be made on whether compliance with the norms should be upheld and, if not, how best to violate the norms. To enable this reasoning to be undertaken, the practical reasoning argumentation scheme with values is used to generate arguments for and against actions such that agents can choose between actions based on their preferences over the values.

The key focus of the paper is on this argument-based account of practical reasoning, which can be used to consider when norms should be violated. The approach is illustrated using a road traffic example that characterises scenarios where the quandary on norm violation may occur. A second and related contribution of the paper is the consideration of what makes an ordering on values acceptable and how such an ordering might be determined.

A final key piece of work from the decade is [Bench-Capon and Atkinson, 2017], which looks at the interaction between dimensions and values. Building on well-established formalisations of factor-based reasoning [Horty, 2017] [Rigoni, 2018], it is shown how values can play several distinct roles in these accounts of legal reasoning, both by explaining preferences between factors and indicating the purposes of the law.

6.2.2 Accrual

Above in Section 5.2.4 early work was described on the accrual of reasons or arguments for the same conclusion. More recently, Gordon's proposed a new model of accrual in a new version of his Carneades system [Gordon and Walton, 2016, Gordon, 2018], with added expressiveness. Then [Prakken, 2019] proposed a new approach in the context of ASPIC+, motivated by some shortcomings in [Prakken, 2005b] and in Gordon's new proposals. [Prakken, 2019] shows that the new proposal satisfies the three proposals of accrual proposed by [Prakken, 2005b] while avoiding the shortcomings of the earlier work.

6.3 Precedential constraint

Early AI and Law work on case-based reasoning was primarily rhetorical in nature in that it was not about computing an 'outcome' or 'winner' of a dispute but instead about generating debates as they can take place between 'good' lawyers. Later work was more logic- and outcome-oriented [Loui et al., 1993, Hage, 1993, Prakken and Sartor, 1998]. This later work inspired a line of research initiated by [Horty, 2011], which aims to formalise the common-law concept of *precedential constraint*, that is, to characterise the conditions under which a decision in a new case is forced or at least allowed by a body of precedents. This is a problem hardly addressed in the initial work on the HYPO and CATO systems. Initially Horty only studied factor-based reasoning but recently he has adapted his approach to dimensions [Horty, 2017, 2019, 2021].

In the factor-based models precedents are simply represented as in CATO, with two sets of factors, respectively, pro and con a boolean decision. Horty's simplest model of precedential constraint is the *result model*, which regards a decision in a new case as *forced* if the precedent cannot be distinguished in the new case, that is, if the new case contains at least all factors pro the decision that the precedent has, while it contains at most all factors con that decision in the precedent. Then a decision is *allowed* if the opposite decision is not forced.

Horty's *reason model* is somewhat more involved. First, it allows to say that in a precedent a subset of the pro decision factors was sufficient for the court to outweigh all the con-decision factors in the case. This subset is called the *rule* of the case. Next, Horty adapts the idea of [Prakken and Sartor, 1998] that a case decision expresses a preference for the pro-decision factors over the con-decision factors in the case. In the following definition, $pro(c)$ and $con(c)$ denote, respectively, the factors pro and con the decision s in case c, $\overline{s}$ is the opposite decision, while $ppro(c)$, which is a subset of $pro(c)$, is the rule of the case. Hence a case decision expresses a preference for any pro-decision set containing at least the pro-decision factors of the case over any con-decision set containing at most the con-decision factors of the case. As in [Prakken and Sartor, 1998], this allows *a fortiori* reasoning from a precedent adding pro-decision factors and/or deleting con-decision factors.

> Let $(ppro(c) \cup con(c), pro(c), s)$ be a case, CB a case base and X and Y sets favouring s and $\overline{s}$, respectively. Then
>
> 1. $Y <_c X$ iff $Y \subseteq con(c)$ and $X \supseteq pro(c)$;
> 2. $Y <_{CB} X$ iff $Y <_c X$ for some $c \in CB$.

Next Horty defines a case base CB to be *inconsistent* if and only if there are factor sets X and Y such that $X <_{CB} Y$ and $Y <_{CB} X$. Then CB is *consistent* if and only if it is not inconsistent. Then Horty defines a decision to be *forced* according to the *reason model* if that is the only way to keep the case base consistent when updated with the new case. Horty proves that for consistent cases bases his result and reason model are equivalent on the assumption that $pro(c) = ppro(c)$ for all cases c.

Quite recently, [van Woerkom et al., 2023] generalised the factor-based result model to deal with hierarchical relations between factors as in CATO's factor hierarchy (see Section 4.1.5), while [Canavotto and Horty, 2023] have done the same for the factor-based reason model. In both cases the definition of precedential constraint is made recursive to allow for reuse of precedents for intermediate decisions.

[Horty, 2019]'s *result model* for *dimension-based* case-based reasoning is quite simple. Cases are now represented as a set of value assignments to dimensions plus a boolean decision. More formally, a *dimension* is a set of values V with two partial orders $\leq_s$ and $\leq_{\overline{s}}$ on V such that $v \leq_s v'$ iff $v' \leq_{\overline{s}} v$. These orderings capture the extent to which different values of a dimension are better for one side and so worse for the other side. Note the difference with [Ashley, 1990]: while Ashley regarded a dimension as always favouring a particular side (although to different degrees), in Horty's approach all that can be said is whether one value favours a side more than another value.

A *value assignment* is a pair (d, v). The functional notation $v(d) = x$ denotes the value x of dimension d. Then a (dimension-based) *case* is a pair $c = (F, outcome(c))$ such that D is a set of dimensions, F is a set of value assignments to all dimensions in D and $outcome(c) \in \{s, \overline{s}\}$. Then a (dimension-based) *case base* is as before a set of cases, but now explicitly assumed to be relative to a set D of dimensions in that all cases assign values to a dimension d iff $d \in D$. Likewise, a (dimension-based) *fact situation* is now an assignment of values to all dimensions in D. As for notation, $v(d, c)$ denotes the value of dimension d in case or fact situation c.

In Horty's *dimension-based result model* of precedential constraint a decision in a fact situation is *forced* iff there exists a precedent c for that decision such that on each dimension the fact situation is at least as favourable for that decision as the precedent. This is formalised with the help of the following preference relation between sets of value assignments.

> Let F and F' be two fact situations with the same set of dimensions. Then $F \leq_s F'$ iff for all $(d, v) \in F$ and all $(d, v') \in F'$ it holds that $v(d) \leq_s v'(d)$.

Then, given a case base CB, deciding fact situation F for s is *forced* iff there exists a case $c = (F', s)$ in CB such that $F' \leq_s F$.

Defining a *dimension-based reason model* of precedential constraint is far more complicated. The main problem is how to define that a subset of the value assignments 'pro' a decision outweighs the value assignments 'con' the decision given that formally value assignments are not categorically pro or con a decision but only better or worse for a decision than other value assignments. In fact, Horty had to revise his initial proposal of [Horty, 2017, 2019] in [Horty, 2021], because of some counterintuitive outcomes of the initial proposal. [Rigoni, 2018] proposes an alternative dimension-based reason model. See [Prakken, 2021] for a formal analysis of these and some other factor- and reason-based models of precedential constraint.

6.4 Use of Arguments

6.4.1 Applications

Numerous example applications have been produced to evaluate evolving computational models of argument, including some scenarios posed by real world problems in an industrial setting. Four such characteristic examples are highlighted here.

The first is a feasibility study conducted in collaboration with a large law firm, and reported in [Al-Abdulkarim et al., 2019], to build a practical system using the ANGELIC methodology [Al-Abdulkarim et al., 2016b] described earlier in this section. In the application, a body of case law relevant for the business was captured as a so-called Abstract Dialectical Framework (ADF) using the ANGELIC methodology. ADFs [Brewka and Woltran, 2010] are a generalisation of Dungian abstract argumentation frameworks. The domain of case law was claims for noise-induced hearing loss against employers. The study involved identification of usable arguments that are key to guiding case handlers in assessing the strength of a negligent hearing loss claim and whether or not it had reasonable prospect of defence. The application of the methodology, and thus the use of ADFs, in this application scenario was shown to be very effective in modelling the domain and providing assistance to case handlers in identifying the arguments relevant for deciding the cases. Subsequently, this line of research was extended [Atkinson et al., 2019] to investigate how the ANGELIC methodology could be used to capture reasoning about factors with magnitude, expanding the range of industrial scenarios that the methodology could be applied in.

A second exemplar real world application setting is given in [Contissa et al., 2013], which uses argument maps to assess liability in the field of air traffic management. In this setting, a 'Legal Case' methodology is used to assist an interdisciplinary team to foresee and mitigate legal problems that may occur through the proposed use of automated technology in air traffic management. The methodology, as described in [Contissa et al., 2013], covers steps to mapping and classifying possible automated technology failures, produce a set of hypotheses of liability link to the failures, and analyse legal rules and arguments supporting the attribution of liability for each of the hypotheses. Although not in full deployment, the argument maps produced are intended for presentation to stakeholders in the domain, including lawyers, to facilitate with the cooperative design and assessment of new technologies for air traffic management.

The third and fourth examples of real-world applications both concern legal argumentation in the domain of law enforcement. [Bex et al., 2016] looked at an AI system for citizen complaint intake about online trade fraud, for example, false web shops or malicious traders on eBay not delivering products to people. In the paper, the first ideas are provided for an argument-based recommender system that, given a complaint form, uses argumentation to determine whether a case is possibly fraud, and then only recommends filing an official report if it is. More information about the intake system can be found in Sections 6.1.3 and 7.3.1. [Testerink et al., 2019] use the same argument-based system, but instead of recommending whether or not to file a report to citizens, it provides responses to messages from international police partners given what is in then police database and certain policy rules.

6.4.2 Policy

Policy making is a domain in which argumentation naturally occurs. Political disputes can turn on disagreements as to objective facts and subjective values, so computational models of argument are well suited to representing these different types of debate. In [Atkinson et al., 2011] a demonstration was given as to how to construct a semantic model on the basis of responses received to a Green Paper, which is a government publication released as part of a consultation process that details specific issues, and then points out possible courses of action in terms of policy and legislation in order to receive feedback from interested parties. An example of the type of debate that has been modelled is an issue in UK Road Traffic policy. The starting situation is that the number of fatal road accidents is an obvious cause for concern on UK roads. There are already speed restrictions in place on various types of road, in the belief that excessive speed causes accidents. The policy issue to be considered is how to reduce road deaths. One option is to introduce speed cameras to discourage speeding. Another is to educate motorists on the dangers of speeding. In [Atkinson et al., 2011] it is shown how a semantic model of this debate can be built from which different policy options can be considered for implementation, representing issues of importance to different stakeholders, such as road safety organisations, motoring lobby groups and civil liberties groups. From the semantic model, arguments for different policy options can be identified with the policy selected being depending up the preferred values being promoted by that option.

6.4.3 Arguments, stories and probabilities in evidential reasoning

Continuing from research done in the 2000s (Section 5.3.3), further work was done on evidential reasoning, enhancing the understanding on how various approaches to the rational handling of the evidence are connected. Three types of normative approaches for the handling of evidence can be distinguished: arguments, scenarios and probabilities [Kaptein et al., 2009, Dawid et al., 2011, Anderson et al., 2005, Verheij et al., 2016, Di Bello and Verheij, 2018]. Each approach can help systematize and regulate how to examine, analyse and weigh the evidence. Where an argument-based approach focuses on dialogue, support and attack, scenario-based approaches highlight explanatory sense-making in coherent, holistic accounts of what has happened, while probabilistic approaches enable the quantitative analysis of evidential strength by connecting to Bayesian modeling and statistics. In the 2010s, the three approaches were studied in various combinations.

One idea was to embed argument structure in Bayesian networks using 'legal idioms' [Fenton et al., 2013, Lagnado et al., 2013, Neil et al., 2019]. The approach aims to systemize the embedding of legally relevant argument structure (for instance about witness reliability and alibi testimony) in Bayesian networks. Inspired by and extending the work on so-called object-oriented Bayesian networks [Hepler et al., 2007], Fenton et al. [2013], Lagnado et al. [2013] propose reusable graphical network structures aimed at the modeling of legal evidential arguments. Idioms are provided for the modeling of evidence accuracy, motive and opportunity, evidential dependencies, alibi evidence and explaining away (downplaying alternative explanations). In its attempt to provide a catalog of reusable argument structure, the idiom approach is similar to work on argumentation schemes [Walton et al., 2008, Walton, 1996], but now in the context of Bayesian network analysis.

This idiom approach was applied to the embedding of evidential stories in Bayesian networks by Vlek et al. [2014, 2016]. In this work, a design method is proposed aimed at alleviating three common difficulties in reasoning with evidence: (1) tunnel vision, (2) the problem of a good story pushing out a true story and (3) finding the relevant variables for a model of the case. The design method uses four idioms: the scenario idiom, representing the events in a story and how they depend on one another (for instance how a burglary developed); the subscenario idiom, allowing for the embedding of one story in another (for instance how the house was entered during a burglary); the variation idiom, used for the modeling of different versions of a story (for instance, entering after smashing a window or picking a lock); and the merged scenarios idiom, used for combining different stories in one Bayesian network model. During the design of a model, four steps are iterated: collecting the relevant scenarios; unfolding scenarios by considering for which story elements evidence can be added; merging the scenarios; and finally adding the evidence. The design method comes with a corresponding explanation format, which allows for the explanation of a Bayesian network model built with the method in terms of the scenarios modeled, the quality of those scenarios (interpreted probabilistically) and the evidential support that is available. The method is evaluated using case studies of real crime cases in the Netherlands.

Work continued on the hybrid integration of argument-based and story-based approaches (Sections 5.1.2 and 5.3.3): [Bex et al., 2010, Bex, 2011] discuss the hybrid theory for stories and arguments about evidence in detail. [Bex and Verheij, 2013, Bench-Capon and Bex, 2015] connect evidential reasoning with stories and arguments to legal reasoning with arguments and cases. [Bex, 2015] further integrates reasoning with stories and arguments, allowing for both of them to be represented and evaluated as elements of a Dung-style argumentation framework [Dung, 1995]. This then allows for different types of reasoning with causality in argumentation frameworks.

There was also work on combining probabilities and arguments in various different ways. [Wieten et al., 2018] discusses an approach for transforming arguments into so-called argument graphs, containing the same kind of causality information as the integrated framework by [Bex, 2015], to Bayesian network structures by using the specific causality information. [Wieten et al., 2019] takes a different approach: arguments are not transformed to Bayesian networks, but they are used in a dialogue to critically analyse Bayesian networks. The paper provides different argument schemes and a dialogue structure for Bayesian network analysis.

[Keppens, 2012] extracts argument diagrams from a Bayesian network so that the evidential reasoning can be scrutinised better by people who do not have in-depth knowledge of Bayesian networks. In a similar vein, [Timmer et al., 2015, 2017] investigates how an argumentation perspective can help in the interpretation of statistical dependency information as modeled by a Bayesian network. For this purpose, support graphs are proposed as an intermediate format. A support graph can disentangle the graphical properties of a Bayesian network and enhances the intuitive interpretation of statistical dependencies. By the use of support graphs, a succinct set of arguments can be generated, reducing superfluous elements.

The case model approach (Section 6.1.4) was originally conceived as a way to connect arguments, scenarios and probabilities in a single modeling approach. The informal ideas of combining the three approaches in one were presented in [Verheij, 2014]. Arguments were intended for addressing the adversarial setting of reasons for and against claims, scenarios for providing a globally coherent perspective, and probabilities for the modeling of gradual uncertainty. The combined approach aimed at keeping the strengths of each of the three separate approaches, while avoiding limitations. For instance, probabilistic approaches provide a well-known and widely useful account of rational evidence handling, they typically require more numerical, statistical information than is reasonably available (in particular Bayesian networks, which model probabilities for all possible combinations of all model variables). Balancing these, the case model approach is presented as 'with and without numbers' by providing an approach that is consistent with a probabilistic analysis but does not require full numerical information. Verheij [2017b] provide the further formal development of the approach. van Leeuwen and Verheij [2019] compares an analysis in terms of Bayesian networks with embedded scenarios and in terms of case models. Both kinds of analysis show how the gradual collection of the evidence has a stepwise influence on how strongly various hypotheses about what has happened are supported in comparison with one another.

Verheij et al. [2016] discuss different combinations of the modeling of evidential reasoning using arguments, scenarios and probabilities, explicating strengths and limitations of each. Prakken et al. [2020] introduce a special issue including various modeling approaches, separately and in hybrid combinations, all using the same real case as an example. Prakken [2020] provides an argumentative analysis, while Fenton et al. [2020] one in terms of Bayesian networks, and Dahlman [2020] applies Bayesian thinking. Koppen and Mackor [2020] use story analysis, and Bex [2020] uses the hybrid argumentative-narrative approach (Section 5.3.3, Figure 11), while Verheij [2020b] gives an analysis with and without probabilities using case models (Section 6.1.4).

6.4.4 Methodology

In [Al-Abdulkarim et al., 2016b] a methodology for capturing case law (ANGELIC) was presented and it was shown how three domains could be modelled using the methodology to capture the factor-based reasoning within those domains and decide cases in accordance with the model. The three domains that were used in the evaluation of the methodology were: the CATO trade secrets cases, cases regarding warrantless search of automobiles, and cases concerning capture and possession of wild animals, which have been popular testbeds in the AI and law literature. The domains are all modelled as ADFs. Once defined for a domain, an ADF can easily be transformed into a logic program that, when instantiated with the facts of a case, can determine outcome for the case and the acceptable arguments leading to this decision. The programs reported in [Al-Abdulkarim et al., 2016b] demonstrated a high degree of success in replicating the outcomes from the cases used in the experiments, yielding a success rate of over 96% accuracy.

The need to maintain the model in the face of evolving case law was recognised in [Al-Abdulkarim et al., 2016a]. There it was shown how the highly modular nature of the ADF facilitated the addition, modification and re-odering of acceptance conditions, as well as the addition and removal of nodes.

6.4.5 Evolution of Case Law

Henderson and Bench-Capon [2019] returned to the topic earlier explored in Section 5.3.2. Again this was based on Levi [1948], and this time focused on changing rules rather than classes. The idea was that each side would present an argument, and the winning argument would form a rule to be applied to future cases.

A number of types of argument were identified:

- *stare decisis*: if a rule covering the current case exists, that rule can be cited as a reason for the decision;
- *class membership*: argues that the current case should be decided in virtue of membership of a particular class; this may broaden or narrow a class used used in an existing rule;
- *floodgate*: argues against a broadening or narrowing on the grounds that it is too big a step and would include or exclude too many cases;
- *exception*: points to a distinguishing feature of a case, and proposes that it should be an exception to the existing rule;
- *logical similarity*: combines two rules with the same outcome into a single rule.

The process is illustrated with three examples: a fictional example based on interpreting the phrase "expected to work", Levi's liability cases beginning with *Dixon v Bell*, and automobile exception to the US 4th Amendment cases involving luggage.

In [Bench-Capon and Henderson, 2019a] the process was made concrete as a set of dialogue moves. The dialogue was conducted over three plies: a proposal, a response and a rebuttal, followed by a judgement resolving the discussion. Each ply was associated with several moves expressing different kinds of argument. An example with sixteen cases is given in [Bench-Capon and Henderson, 2019b].

6.4.6 Statutory Interpretation

While most work on argumentation in AI and law concerns reasoning with legal cases, argumentation also has a role to play in the interpretation of statutes. Often the interpretation of a term in a statute is not clear: should it be given a literal interpretation, or interpreted according to its context and the purpose of the statute?

In [Sartor et al., 2014] two jurisprudential sources are used to identify the kinds of arguments that can be used. MacCormick and Summers [2016] identify eleven types of arguments and Tarello [1980] identifies fourteen. Tarello's list complements MacCormick and Summers' list, since the latter focuses on the kinds of input on which the interpretive argument is based (ordinary language, technical language, statutory context, precedent, etc.) while the first focuses on the reasoning steps by which the interpretive argument is constituted. Where conflicting arguments of different types are available, criteria are needed: a list of such criteria is given in [Alexy and Dreier, 1991]. This jurisprudential work is used to provide a general logical structure for arguments based on interpretative canons, and for arguments about which should be followed in cases of conflict. This approach was formalised in defeasible deontic logic with canons taken as defeasible rules in [Rotolo et al., 2015].

In [Walton et al., 2016] the canons of MacCormick and Summers and Tarello were presented as argumentation schemes, some positive arguing for an interpretation and some negative arguing against an interpretation. Counterargments can be generated using the three associated critical questions. It is illustrated with a case modelled using Carneades [Gordon, 2013].

7 Current developments: The breakthrough of AI in society

Currently, AI has become a wide-spread topic of discussion throughout society. Both its risks and its limitations are addressed in public debate and in research. Machine prediction of decisions is studied, both as a tool and as a risk. Notwithstanding widespread efforts, the aligning of learning and reasoning remains a research challenge. Large language models (in particular since ChatGPT's public release in November 2022) are used by virtually everyone. Ethical concerns are discussed, by philosophers, tech developers and AI researchers. Machine learning is now also 'good-old fashioned', a term before only applied to symbolic AI methods. Attempts are made to arrive at new hybrid AI approaches connecting knowledge, reasoning, learning and language. Steps are made to use the argumentation approach as a model of such integration (cf. [Verheij, 2020a]).

7.1 Argument Generation

7.1.1 Argumentation Schemes

In [Atkinson and Bench-Capon, 2021] a summary is given of the impact of Walton's argumentation schemes on research in AI and Law research. Within that discussion it is shown how the systematisation of natural patterns of arguments can be done into schemes to enable arguments to be generated from these schemes. There are a number of ways in which schemes can be encoded. Logic programming can be used straighforwardly to represent the schemes as rules, or they could be captured as defeasible rules within a framework such as ASPIC+ [Prakken, 2010]. Hand coding is required to encode schemes as rules in this way, but other works have taken a more general approach by building inference engines to execute the schemes. One of the richest tools developed to meet this aim is the Carneades system [Gordon, 2013]. The system is an integrated set of tools for argument (re)construction, evaluation, mapping and interchange. Carneades provides a library of 106 schemes but with the ability to extend this list with the specification of additional schemes. Utilising these schemes, and their critical questions, allows for the generation of arguments and counter arguments. Carneades also allows for the evaluation of arguments using several different standards of proof required for acceptance of a given argument.

7.1.2 Rationales

Building on work in the 1990s on rationales and argument moves [Loui and Norman, 1995] (discussed in Section 4.1.2), Bench-Capon and Verheij [2022] show how methods of computational argumentation developed later can be applied to the unpacking of arguments. Concretely, examples of a compression rationale and of a resolution rationale were analyzed. The compression rationale example involves the unpacking of an intermediate step (as discussed in Section 4.1.2, Figure 5) and that of a resolution rationale the unpacking of an implicit conflict resolution involving a preference. Methods applied include structured argumentation approach (similar to ASPIC+ [Prakken, 2010]) and a sentence-based approach (DefLog [Verheij, 2003b]). It is concluded that modern approaches can make the unpackings explicit, but do not retain the formal connection between the unpacked and the unpacking arguments.

7.2 Evaluation of Arguments

Whilst teleological reasoning is a well established feature of AI and law research, as discussed within this chapter, new models for reasoning about values continue to be developed. In [Maranhão et al., 2021] an additive model of balancing values is set out whereby factors intensify or attenuate impacts on values and values are assigned degrees of relative importance. What results is an assessment of an action's impacts on single values, which are then aggregated to determine the action's total impact on the sets of values promoted and demoted. Comparing an action's impacts on the promoted and demoted values then yields a determination as to whether the action is either permitted or prohibited.

Supplementing the balancing model are formal definitions of change functions that induce shifts in the balance of values through addition or subtraction of factors, or by additions or subtractions of values in the model. These operations are intended to have some resemblance to argument moves, where new features of the legal case or moral considerations are brought into play to oppose previously justified conclusions.

7.3 Use of Arguments

7.3.1 Applications in law enforcement

In the 2020's, the ideas and prototypes of argument-based applications for law enforcement (cf. Sections 6.1.3. 6.4.1) are implemented and used at scale at the Dutch Police. The intake and analysis of citizen crime reports regarding online trade fraud are handled by an online recommender system that uses argumentation ([Odekerken et al., 2020, 2022]). Furthermore, [Odekerken and Bex, 2020] adapt the case-based reasoning model of precedential constraint (see Section 6.3) for the classification of possibly fraudulent webshops, extending the model of [Horty, 2011] to also deal with incomplete cases and inconsistent case bases [Odekerken et al., 2023a,b].

7.3.2 Machine learning and explanations

Following the earlier work of [Ashley and Brüninghaus, 2009, Ashley and Walker, 2013] and [Schraagen and Bex, 2018], further research was done into models that extract basic facts or factors from text using data driven NLP methods and then reason with these fact(ors) using logical models of argumentation and case-based reasoning. [Mumford et al., 2021] use a BERT model to extract factors from cases of the ECHR, after which they reason with these cases using logical models of argument (cf. Section 7.3.3).

[Prakken and Ratsma, 2022] provide an argumentation method with which the outcome of black box (machine learning) systems can be explained post-hoc. Based on the model of precedential constraint (6.3), they define a dialogue-based argumentation model in which the proponent cites the case that is to be explained (i.e. why did case c have outcome o?), and the opponent then tries to defeat this cited case by arguing, for example, that it has missing or additional factors that might influence the outcome o. The model was later extended by [Peters et al., 2023] so that it could reason with inconsistent case bases and also have the case bases be constructed by the black box model.

Steging et al. [2021a,b] address the issue that machine learning systems can draw the right conclusions for the wrong reasons, in the sense that high accuracies do not imply that the correct conditions are used in a trained model. Whereas in image classification, such a mismatch between reasons and conclusion may not be problematic, the use of unsound rationales is unwanted in legal applications where the justification of a decision is of central relevance.

Steging et al. [2021a] develops a human-in-the-loop approach to investigate and improve the rationale used by machine learning models.

The method is hybrid in two ways. First it is an example of a hybrid intelligence approach [Akata et al., 2020] in which humans and machines augment each others' performance. In this case, the human knowledge that is available (although perhaps incomplete) can be used for improvement of the rationale used by a machine learning model. Second the approach is hybrid by combining different methods in AI, in particular by the use of both machine learning and knowledge representation methods. Steging et al. [2021b] applies explainable AI methods that detect the features used for decision making in a trained model. The paper shows that even with high accuracy and good relevant feature detection, the use of a correct rationale is not guaranteed.

In this research [Steging et al., 2021a,b], synthetic data sets with a known structure are used. The data sets are generated using a given knowledge structure, so that a correct ground truth rationale is known beforehand and can be used for method evaluation. One data set concerns a fictional welfare benefit domain about eligibility of a person for a welfare benefit to cover the expenses for visiting their spouse in the hospital. The domain and data set were introduced by Bench-Capon [1993] in order to investigate whether a neural network can correctly learn a rule from data. The other data set models actual Dutch tort law based on the articles 6:162 and 6:163 of the Dutch civil code about when an unlawful act legally determines a duty to repair the damages caused (cf. also Figure 12). The data sets are publicly available [Steging et al., 2023].

7.3.3 Methodology

The first full account of the ANGELIC methodology was set out in 2016 [Al-Abdulkarim et al., 2016b] and since then it has been extended and applied to a range of legal domain scenarios, most recently the popular domain of the European Convention on Human Rights. In [Collenette et al., 2023] it was shown how Article 6 of the convention, covering the right to a fair trial, could be modelled using the ANGELIC methodology. The model was then evaluated using forty cases heard in the European Court of Human Rights (ECHR) to determine whether the ANGELIC model could produce the same outcomes that the judges had in the original cases. A 97% success rate was reported with this exercise and, crucially, the program was able to give easily digestible explanations as to why it had arrived at its outcome of whether or not there had been a violation of Article 6 in each of the cases under consideration. The current version of the ANGELIC methodology is presented in [Atkinson and Bench-Capon, 2023].

Despite the success in terms of both accuracy and explainability, there were still parts of the process of constructing the domain model that rely on manual analysis, specifically the ascription of factors from cases to the model. To automate this task within the overall process, a model was proposed in [Mumford et al., 2022] to use machine learning for the factor ascription task, such that once factors present in a case are ascribed, the outcome follows from reasoning over the domain model. The approach yields a hybrid AI model combining symbolic and data-driven approaches. The most recent line of work reported on a study involving the annotation of a corpus of Article 6 cases, yielding insight on the distribution of the factors relevant to the complaint of a potential violation of Article 6. The study produced an annotated data set for training models, using natural language processing techniques, to perform the factor ascription task in accordance with the ADF for Article 6. Encouraging results were reported from experiments on this task, providing impetus for further exploring the hybrid use of AI techniques for supporting automated reasoning about legal cases.

7.3.4 Statutory Interpretation

Statutory reasoning was discussed in Section 6.4.6: a different approach was proposed in [Araszkiewicz et al., 2020]. This paper introduces the notion of reasoning protocol as a frame for a set of elements used by relevant agents to justify their claims. The model allows the representation of reasoning using not only factors, but also about the relevance of factors in deciding legal cases on the basis of statutory rules. After defining the various elements of the protocol, the paper investigates selected patterns of case-based judicial reasoning in the context of statutory interpretation as understood in continental legal culture.

A second paper by Araszkiewicz [2022] takes the work described in Section 6.4.6 as its starting point and extends it with a layer of case base reasoning reasoning with and about default preference relations between (classes of) interpretive canons. A set of factors supporting preference for linguistic canons over teleological canons and *vice versa* are identified. These are then used with rules extracted from precedent cases the manner of [Prakken and Sartor, 1998]. An argumentation scheme to represent the reasoning is provided.

Concluding remarks

This chapter has showcased the close interaction between research in the theory of computational argumentation and the field of AI and Law. By exploring the early days and historical developments decade by decade, a natural continuity of mutual inspiration between the fields has been presented.

The work described over the timeline has covered the generation, evaluation and use of arguments in AI and Law. In the earlier years, many of the approaches were of a semi-formal nature, then these were followed by the development of rule-based and case-based approaches that mirrored developments in the field of general AI. These approaches were then overtaken by significant advances in topics on computational argumentation, which developed into a field in its own right and brought forth a much more formal approach to modelling legal reasoning. The review of developments closed with coverage of how data-driven approaches to AI that have received significant attention in recent times are being brought to bear on tasks involving the modelling of arguments in legal settings.

In addition to the continued development of specific techniques for argument-based approaches to AI and law, more research is emerging demonstrating the integration of knowledge-based and data-driven approaches, with the aim of producing hybrid systems that seek to reap the benefits of these distinct approaches. With argumentation playing such a strong role in human-based legal reasoning, we can expect to see computational models of argument remaining of importance for driving forward research in AI and law, and leading to applications in the legal domain that contribute to important aims within the topic of explainable and trustworthy AI.

References

Z. Akata, D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerincx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen, and M. Welling. A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(08):18–28, August 2020.

L. Al-Abdulkarim, K. Atkinson, and T. Bench-Capon. Accommodating change. *Artificial Intelligence and Law*, 24:409–427, 2016a.

L. Al-Abdulkarim, K. Atkinson, and T. Bench-Capon. A methodology for designing systems to reason with legal cases using Abstract Dialectical Frameworks. *Artificial Intelligence and Law*, 24:1–49, 2016b.

L. Al-Abdulkarim, K. Atkinson, T. Bench-Capon, S. Whittle, R. Williams, and C. Wolfenden. Noise induced hearing loss: Building an application using the ANGELIC methodology. *Argument and Computation*, 10(1):5–22, 2019.

V. Aleven. *Teaching case-based argumentation through a model and examples*. Ph.D. thesis, University of Pittsburgh, 1997.

V. Aleven. Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence*, 150(1-2):183–237, 2003.

V. Aleven and K. D. Ashley. Doing things with factors. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 31–41. ACM Press, 1995.

V. Aleven and K. D. Ashley. Evaluating a learning environment for case-based argumentation skills. In *Proceedings of the 6th International Conference on Artificial Intelligence and Law*, pages 170–179. ACM Press, 1997.

R. Alexy. *Theorie der juristischen Argumentation. Die Theorie des rationalen Diskurses als eine Theorie der juristischen Begründung.* Suhrkamp Verlag, Frankfurt am Main, 1978.

R. Alexy and R. Dreier. Statutory interpretation in the Federal Republic of Germany. In D. MacCormick and R. Summers, editors, *Interpreting Statutes: A Comparative Study,*, pages 72–101. Dartmouth Aldershot/Brookfield USA/Hong Kong/Singapore/Sydney, 1991.

L. Amgoud and C. Cayrol. A model of reasoning based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–215, 2002.

T. Anderson, D. Schum, and W. Twining. *Analysis of Evidence. 2nd Edition.* Cambridge University Press, Cambridge, 2005.

G. Antoniou, D. Billington, G. Governatori, and M. Maher. A flexible framework for defeasible logics. In *Proceedings of the 17th National Conference on Artificial Intelligence*, pages 405–410, 2000.

M. Araszkiewicz. A hybrid model of argument concerning preferences between statutory interpretation canons. In E. Francesconi, G. Borges, and C. Sorge, editors, *Legal Knowledge and Information Systems. JURIX 2022: The Twenty-fifth Annual Conference*, pages 3–12. IOS Press, Amsterdam, 2022.

M. Araszkiewicz, T. Żurek, and B. Kuźniacki. Reasoning with and about factors in statutory interpretation. In *Proceedings of the 4th International Workshop on MIning and REasoning with Legal texts. CEUR Workshop Proceedings*, volume 2632, 2020.

A. Artikis, M. Sergot, and J. Pitt. An executable specification of an argumentation protocol. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 1–11. ACM Press, 2003.

K. Ashley. A brief history of the changing roles of case prediction in AI and law. *Law in Context*, 36:93–112, 2019.

K. Ashley and S. Brüninghaus. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17(2):125–165, 2009.

K. D. Ashley. Toward a computational theory of arguing with precedents. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, pages 93–102. ACM Press, 1989.

K. D. Ashley. *Modeling legal arguments: Reasoning with cases and hypotheticals*. MIT press, Cambridge, Mass., 1990.

K. D. Ashley. Reasoning with cases and hypotheticals in HYPO. *International Journal of Man-Machine Studies*, 34(6):753–796, 1991.

K. D. Ashley and V. R. Walker. Toward constructing evidence-based legal arguments using legal decision documents and machine learning. In *Proceedings of the 14th International Conference on Artificial Intelligence and Law*, pages 176–180. ACM Press, 2013.

K. D. Ashley, N. Pinkwart, C. Lynch, and V. Aleven. Learning by diagramming supreme court oral arguments. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 271–275. ACM Press, 2007.

K. Atkinson and T. Bench-Capon. Legal case-based reasoning as practical reasoning. *Artificial Intelligence and Law*, 13:93–131, 2005.

K. Atkinson and T. Bench-Capon. Practical reasoning as presumptive argumentation using Action based Alternating Transition Systems. *Artificial Intelligence*, 171:855–874, 2007.

K. Atkinson and T. Bench-Capon. Argumentation schemes in AI and Law. *Argument and Computation*, 12(3):417–434, 2021.

K. Atkinson and T. Bench-Capon. ANGELIC II: An improved methodology for representing legal domain knowledge. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pages 12–21. ACM Press, 2023.

K. Atkinson, T. Bench-Capon, and P. McBurney. Arguing about cases as practical reasoning. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 35–44. ACM Press, 2005.

K. Atkinson, T. Bench-Capon, and P. McBurney. Computational representation of persuasive argument. *Synthese*, 152:157–206, 2006a.

K. Atkinson, T. Bench-Capon, and P. McBurney. Parmenides: facilitating deliberation in democracies. *Artificial Intelligence and Law*, 14: 261–275, 2006b.

K. Atkinson, T. Bench-Capon, D. Cartwright, and A. Wyner. Semantic models for policy deliberation. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, pages 81–90. ACM Press, 2011.

K. Atkinson, T. Bench-Capon, H. Prakken, and A. Wyner. Argumentation schemes for reasoning about factors with dimensions. In K. Ashley, editor, *Legal Knowledge and Information Systems. JURIX 2013: The Twenty-sixth Annual Conference*, pages 39–48. IOS Press, Amsterdam, 2013.

K. Atkinson, T. Bench-Capon, T. Routen, A. Sánchez, S. Whittle, R. Williams, and C. Wolfenden. Realising ANGELIC designs using logiak. In M. Araszkiewicz and V. Rodríguez-Doncel, editors, *Legal Knowledge and Information Systems. JURIX 2019: The Thirty-second Annual Conference*, pages 151–156. IOS Press, 2019.

T. Bench-Capon. *Knowledge Representation: An Approach to Artificial Intelligence*. Academic Press, 1990.

T. Bench-Capon. Neural networks and open texture. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 292–297. ACM Press, 1993.

T. Bench-Capon. Arguing with cases. In A. Oskamp, R. de Mulder, C. V. Noortwijk, C. Grütters, K. Ashley, and T. Gordon, editors, *Legal Knowledge-Based Systems. JURIX: The Tenth Conference*, pages 85–100. IOS Press, 1997.

T. Bench-Capon. Specification and implementation of Toulmin dialogue game. In J. Hage, T. Bench-Capon, A. Koers, C. de Vey Mestdagh, and C. Grütters, editors, *Legal Knowledge-Based Systems. JURIX: The Eleventh Conference*, pages 5–19, Nijmegen, 1998. Gerard Noodt Instituut.

T. Bench-Capon. Some observations on modelling case based reasoning with formal argument models. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law*, pages 36–42. ACM Press, 1999.

T. Bench-Capon. The missing link revisited: The role of teleology in representing legal argument. *Artificial Intelligence and Law*, 10:79, 2002a.

T. Bench-Capon. Representation of case law as an argumentation framework. In T. Bench-Capon, A. Daskalopulu, and R. Winkels, editors, *Legal Knowledge and Information Systems. JURIX 2002: The Fifteenth Annual Conference*, pages 53–62, Amsterdam etc, 2002b. IOS Press.

T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13: 429–448, 2003.

T. Bench-Capon. HYPO's legacy: introduction to the virtual special issue. *Artificial Intelligence and Law*, 25:205–250, 2017.

T. Bench-Capon. Before and after Dung: Argumentation in AI and Law. *Argument and Computation*, 11(1-2):221–238, 2020.

T. Bench-Capon and K. Atkinson. Dimensions and values for legal CBR. In A. Wyner and G. Casini, editors, *Legal Knowledge and Information Systems. JURIX 2017: The Thirtieth Annual Conference*, pages 27–32, Amsterdam, 2017. IOS Press.

T. Bench-Capon and K. Atkinson. Precedential constraint: The role of issues. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 12–21. ACM Press, 2021.

T. Bench-Capon and F. Bex. Cases and stories, dimensions and scripts. In A. Rotolo, editor, *Legal Knowledge and Information Systems. JURIX 2015: The Twenty-Eight Annual Conference*, pages 11–20. IOS Press, 2015.

T. Bench-Capon and T. Gordon. Implementing a theory of a legal domain. In E. Francesconi, G. Borges, and C. Sorge, editors, *Legal Knowledge and Information Systems. JURIX 2022: The Thirty-Fifth Annual Conference*, pages 13–22. IOS Press, Amsterdam, 2022.

T. Bench-Capon and J. Henderson. A dialogical model of case law dynamics. In M. Araszkiewicz and V. Rodriguez-Doncel, editors, *Legal Knowledge and Information Systems. JURIX 2019: The Thirty-Second Annual Conference*, pages 163–168. IOS Press, Amsterdam etc., 2019a.

T. Bench-Capon and J. Henderson. Modelling case law dynamics with dialogue moves. Technical report, Technical Report ULCS-19-003, University of Liverpool, 2019b.

T. Bench-Capon and S. Modgil. Norms and value based reasoning: justifying compliance and violation. *Artificial Intelligence and Law*, 25 (1):29–64, 2017.

T. Bench-Capon and H. Prakken. Using argument schemes for hypothetical reasoning in law. *Artificial Intelligence and Law*, 18:153–174, 2010.

T. Bench-Capon and E. L. Rissland. Back to the future: dimensions revisited. In B. Verheij, A. Lodder, R. Loui, and A. Muntjewerff, editors, *Legal Knowledge and Information Systems. JURIX 2001: The Fourteenth Annual Conference*, pages 41–52, Amsterdam etc, 2001. IOS Press.

T. Bench-Capon and G. Sartor. Theory based explanation of case law domains. In *Proceedings of the 8th International Conference on Artificial Intelligence and Law*, pages 12–21. ACM Press, 2001.

T. Bench-Capon and G. Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150(1-2):97–143, 2003.

T. Bench-Capon and M. Sergot. Towards a rule-based representation of open texture in law. In C. Walter, editor, *Computer Power and Legal Language*, pages 39–61. Quorum Books: New York, 1988.

T. Bench-Capon and G. Staniford. Plaid: proactive legal assistance. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 81–88. ACM Press, 1995.

T. Bench-Capon and B. Verheij. Unpacking arguments. In E. Francesconi, G. Borges, and C. Sorge, editors, *Legal Knowledge and Information Systems. JURIX 2022: The Twenty-fifth Annual Conference*, pages 145–150, Amsterdam, 2022. IOS Press.

T. Bench-Capon, P. Dunne, and P. Leng. Interacting with knowledge-based systems through dialogue games. In *Proceedings of the Eleventh International Conference on Expert Systems and Applications*, pages 123–140, 1991.

T. Bench-Capon, T. Geldard, and P. H. Leng. A method for the computational modelling of dialectical argument with dialogue games. *Artificial Intelligence and Law*, 8:233–254, 2000.

T. Bench-Capon, H. Prakken, and G. Sartor. Argumentation in legal reasoning. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, pages 363–382. Springer, Berlin, 2009.

T. Bench-Capon, H. Prakken, A. Wyner, and K. Atkinson. Argument schemes for reasoning with legal cases using values. In *Proceedings of the 14th International Conference on Artificial Intelligence and Law*, pages 13–22. ACM Press, 2013.

T. Bench-Capon, K. Atkinson, and A. Wyner. Using argumentation to structure e-participation in policy making. *Trans. Large Scale Data Knowl. Centered Syst.*, 18:1–29, 2015.

D. H. Berman and C. D. Hafner. Incorporating procedural context into a model of case-based legal reasoning. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*, pages 12–20. ACM Press, 1991.

D. H. Berman and C. D. Hafner. Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 50–59. ACM Press, 1993.

D. H. Berman and C. D. Hafner. Understanding precedents in a temporal context of evolving legal doctrine. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, pages 42–51. ACM Press, 1995.

F. Bex. Analysing stories using schemes. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, pages 93–116. Ashgate Publishing, Farnham, 2009.

F. Bex. *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Springer, Berlin, 2011.

F. Bex. An integrated theory of causal stories and evidential arguments. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 13–22. ACM Press, 2015.

F. Bex. The hybrid theory of stories and arguments applied to the Simonshaven case. *TopiCS in Cognitive Science*, 12(4):1152–1174, 2020.

F. Bex and H. Prakken. Reinterpreting arguments in dialogue: an application to evidential reasoning. In T. Gordon, editor, *Legal Knowledge and Information Systems: JURIX 2004, the Seventeenth Annual Conference*, pages 119–129. IOS Press, 2004.

F. Bex and B. Verheij. Legal stories and the process of proof. *Artificial Intelligence and Law*, 21:253–278, 2013.

F. Bex, H. Prakken, C. Reed, and D. Walton. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, 11:125, 2003.

F. Bex, H. Prakken, and B. Verheij. Anchored narratives in reasoning about evidence. In T. M. v. Engers, editor, *Legal Knowledge and Information Systems: JURIX 2006: the Nineteenth Annual Conference*, volume 152, page 11. IOS Press, 2006.

F. Bex, H. Prakken, and B. Verheij. Formalising argumentative story-based analysis of evidence. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 1–10. ACM Press, 2007a.

F. Bex, S. Van den Braak, H. Van Oostendorp, H. Prakken, B. Verheij, and G. Vreeswijk. Sense-making software for crime investigation: how to combine stories and arguments? *Law, Probability and Risk*, 6(1-4): 145–168, 2007b.

F. Bex, T. Bench-Capon, and K. Atkinson. Did he jump or was he pushed? Abductive practical reasoning. *Artificial Intelligence and Law*, 17:79–99, 2009.

F. Bex, P. J. Van Koppen, H. Prakken, and B. Verheij. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law*, 18:123–152, 2010.

F. Bex, J. Peters, and B. Testerink. AI for online criminal complaints: From natural dialogues to structured scenarios. In *Artificial intelligence for justice workshop (ECAI 2016)*, 2016.

L. Bodenstaff, H. Prakken, and G. Vreeswijk. On formalising dialogue systems for argumentation in the event calculus. In *Proceedings of the Eleventh International Workshop on Nonmonotonic Reasoning*, pages 374–382, Windermere, UK, 2006.

A. Bondarenko, P. Dung, R. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

L. Branting. *Reasoning with Rules and Precedents: A Computational Model of Legal Analysis.* Kluwer Academic Publishers, Dordrecht/-Boston/London, 1999.

G. Brewka. Dynamic argument systems: a formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation*, 11:257–282, 2001.

G. Brewka and T. Gordon. Carneades and Abstract Dialectical Frameworks: A reconstruction. In P. Baroni, F. Cerutti, M. Giacomin, and G. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 3–12. IOS Press, Amsterdam, 2010.

G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, pages 102–111. AAAI Press, 2010.

S. Brüninghaus and K. Ashley. Predicting outcomes of case based legal arguments. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 233–242. ACM Press, 2003.

B. G. Buchanan and E. H. Shortliffe. *Rule based expert systems: the Mycin experiments of the Stanford heuristic programming project.* Addison-Wesley Longman Publishing Co., Inc., 1984.

R. Calegari, R. Riveret, and G. Sartor. The burden of persuasion in structured argumentation. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 180–184. ACM Press, 2021.

I. Canavotto and J. Horty. Reasoning with hierarchies of open-textured predicates. In *Proceedings of the 19th International Conference on Artificial Intelligence and Law*, pages 52–61. ACM Press, 2023.

A. Chorley and T. Bench-Capon. Agatha: Using heuristic search to automate the construction of case law theories. *Artificial Intelligence and Law*, 13(1):9–51, 2005a.

A. Chorley and T. Bench-Capon. An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law*, 13(3):323–371, 2005b.

P. Clark and T. Niblett. The CN2 induction algorithm. *Machine Learning*, 3:261–283, 1989.

J. Collenette, K. Atkinson, and T. Bench-Capon. Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights. *Artificial Intelligence*, page 103861, 2023.

G. Contissa, M. Laukyte, G. Sartor, and H. Schebesta. Assessing liability with argumentation maps: An application in aviation law. In K. Ashley, editor, *Legal Knowledge and Information Systems. JURIX 2013: The Twenty-sixth Annual Conference*, pages 73–76, Amsterdam, 2013. IOS Press.

C. Dahlman. De-biasing legal fact-finders with bayesian thinking. *TopiCS in Cognitive Science*, 12(4):1115–1131, 2020.

A. P. Dawid, W. Twining, and M. Vasiliki, editors. *Evidence, Inference and Enquiry.* Oxford University Press, Oxford, 2011.

M. Di Bello and B. Verheij. Evidential reasoning. In G. Bongiovanni, G. Postema, A. Rotolo, G. Sartor, C. Valentini, and D. N. Walton, editors, *Handbook of Legal Reasoning and Argumentation*, pages 447–493. Springer, Dordrecht, 2018.

J. P. Dick. Representation of legal text for conceptual retrieval. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Law*, pages 244–253. ACM Press, 1991.

P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, volume 93, pages 852–857, 1993a.

P. M. Dung. An argumentation semantics for logic programming with explicit negation. In *Proceedings of the Tenth Logic Programming Conference*, pages 616–630, Cambridge, MA, 1993b. MIT Press.

P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.

R. Dworkin. *Taking rights seriously.* Duckworth, London, 1978.

N. E. Fenton, M. D. Neil, and D. A. Lagnado. A general structure for legal arguments about evidence using Bayesian Networks. *Cognitive Science*, 37:61–102, 2013.

N. E. Fenton, M. D. Neil, B. Yet, and D. A. Lagnado. Analyzing the Simonshaven case using Bayesian networks. *TopiCS in Cognitive Science*, 12(4):1092–1114, 2020.

K. Freeman and A. M. Farley. A model of argumentation and its application to legal reasoning. *Artificial Intelligence and Law*, 4:163–197, 1996.

D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Nonmonotonic Reasoning and Uncertain Reasoning.* Clarendon Press, Oxford, 1994.

H. Geffner and J. Pearl. Conditional entailment: bridging two approaches to default reasoning. *Artificial Intelligence*, 53:209–244, 1992.

E. Giunchiglia, J. Lee, V. Lifschitz, N. McCain, and H. Turner. Nonmonotonic causal theories. *Artificial Intelligence*, 153:49–104, 2004.

T. Gordon. Defining argument weighing functions. *Journal of Applied Logics – IfCoLog Journal of Logics and their Application*, 5:747–773, 2018.

T. Gordon and D. Walton. Formalizing balancing arguments. In P. Baroni, T. Gordon, T. Scheffler, and M. Stede, editors, *Computational Models of Argument. Proceedings of COMMA 2016*, pages 327–338. IOS Press, Amsterdam, 2016.

T. F. Gordon. The pleadings game: Formalizing procedural justice. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 10–19. ACM Press, 1993a.

T. F. Gordon. The pleadings game. *Artificial Intelligence and Law*, 2 (4):239–292, 1993b.

T. F. Gordon. Combining rules and ontologies with Carneades. In S. B. et al., editor, *Proceedings of the 5th International RuleML2011@BRF Challenge, at the 5th International Web Rule Symposium*, volume 799, pages 103–110. CEUR Workshop Proceedings, 2011.

T. F. Gordon. Introducing the Carneades web application. In *Proceedings of the 14th International Conference on Artificial Intelligence and Law*, pages 243–244, 2013.

T. F. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the 6th International Conference on Artificial Intelligence and Law*, pages 10–18. ACM Press, 1997.

T. F. Gordon and D. Walton. Legal reasoning with argumentation schemes. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 137–146. ACM Press, 2009.

T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

G. Governatori and G. Sartor. Burdens of proof in monological argumentation. In R. Winkels, editor, *Legal Knowledge and Information Systems. JURIX 2010: The Twenty-Third Annual Conference*, pages 37–46. IOS Press, Amsterdam etc., 2010.

G. Governatori, A. Rotolo, and G. Sartor. Temporalised normative positions in Defeasible Logic. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 25–34. ACM Press, 2005.

G. Governatori, A. Rotolo, R. Riveret, M. Palmirani, and G. Sartor. Variants of temporal defeasible logics for modelling norm modifications. In *Proceedings of 11th International Conference on Artificial Intelligence and Law*, pages 155–9. ACM Press, 2007.

G. Governatori, F. Olivieri, A. Rotolo, and S. Scannapieco. Computing strong and weak permissions in Defeasible Logic. *Journal of Philosophical Logic*, 42:799–829, 2013.

G. Governatori, B. Verheij, M. Araszkiewicz, T. Bench-Capon, E. Francesconi, and M. Grabmair. Thirty years of AI and Law: The first decade. *Artificial Intelligence and Law*, 30(4):481–519, 2022.

M. Grabmair. *Modeling Purposive Legal Argumentation and Case Outcome Prediction using Argument Schemes in the Value Judgment Formalism*. PhD thesis, University of Pittsburgh, 2016.

M. Grabmair. Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs. In *Proceedings of the 16th International Conference on Artificial Intelligence and Law*, pages 89–98. ACM Press, 2017.

M. Grabmair and K. Ashley. Facilitating case comparison using value judgments and intermediate legal concepts. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, pages 161–170. ACM Press, 2011.

M. Grabmair and K. D. Ashley. Argumentation with value judgments-an example of hypothetical reasoning. In R. Winkels, editor, *Legal Knowledge and Information Systems. JURIX 2010: The Twenty-Third Annual Conference*, pages 67–76. IOS Press, Amsterdam, 2010.

C. Grover, B. Hachey, I. Hughson, and C. Korycinski. Automatic summarisation of legal documents. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 243–251. ACM Press, 2003.

B. Hachey and C. Grover. Automatic legal text summarisation: experiments with summary structuring. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 75–84. ACM Press, 2005.

B. Hachey and C. Grover. Extractive summarisation of legal texts. *Artificial Intelligence and Law*, 14:305–345, 2006.

C. D. Hafner and D. H. Berman. The role of context in case-based legal reasoning: teleological, temporal, and procedural. *Artificial Intelligence and Law*, 10(1-3):19–64, 2002.

J. C. Hage. Monological reason-based logic: A low level integration of rule-based reasoning and case-based reasoning. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 30–39. ACM Press, 1993.

J. C. Hage. A theory of legal reasoning and a logic to match. *Artificial Intelligence and Law*, 4(2):199–273, 1996.

J. C. Hage, R. Leenes, and A. Lodder. Hard cases: a procedural approach. *Artificial Intelligence and Law*, 2:113–167, 1993.

C. L. Hamblin. *Fallacies.* Methuen, London, 1970.

J. Henderson and T. Bench-Capon. Dynamic arguments in a case law domain. In *Proceedings of the 8th International Conference on Artificial Intelligence and Law*, pages 60–69. ACM Press, 2001.

J. Henderson and T. Bench-Capon. Describing the development of case law. In *Proceedings of the 17th International Conference on Artificial Intelligence and Law*, pages 32–41. ACM Press, 2019.

A. B. Hepler, A. P. Dawid, and V. Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6(1–4):275–293, 2007.

J. Horty. Rules and reasons in the theory of precedent. *Legal Theory*, 17:1–33, 2011.

J. Horty. Reasoning with dimensions and magnitudes. In *Proceedings of the 16th International Conference on Artificial Intelligence and Law*, pages 109–118. ACM Press, 2017.

J. Horty. Reasoning with dimensions and magnitudes. *Artificial Intelligence and Law*, 27:309–345, 2019.

J. Horty. Modifying the reason model. *Artificial Intelligence and Law*, 29:271–285, 2021.

J. F. Horty and T. Bench-Capon. A factor-based definition of precedential constraint. *Artificial Intelligence and Law*, 20(2):181–214, 2012.

H. Jakobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law*, pages 53–62. ACM Press, 1999.

J. R. Josephson and S. G. Josephson. *Abductive inference: Computation, philosophy, technology.* Cambridge University Press, 1996.

H. Kaptein, H. Prakken, and B. Verheij, editors. *Legal Evidence and Proof: Statistics, Stories, Logic (Applied Legal Philosophy Series).* Ashgate, Farnham, 2009.

J. Keppens. Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence and Law*, 20:109–143, 2012.

J. Keppens and J. Zeleznikow. A model based reasoning approach for generating plausible crime scenarios from evidence. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 51–59. ACM Press, 2003.

J. L. Kolodner, R. L. Simpson, and K. Sycara-Cyranski. *A process model of cased-based reasoning in problem solving.* School of Information and Computer Science, Georgia Institute of Technology, 1985.

P. Koppen and A. Mackor. A scenario approach to the Simonshaven case. *TopiCS in Cognitive Science*, 12(4):1132–1151, 2020.

R. A. Kowalski and F. Toni. Abstract argumentation. *Artificial Intelligence and Law*, 4(3-4):275–296, 1996.

D. A. Lagnado, N. Fenton, and M. Neil. Legal idioms: a framework for evidential reasoning. *Argument and Computation*, 4(1):46–63, 2013.

I. Lakatos. *Proofs and refutations.* Nelson London, 1963.

E. H. Levi. *An Introduction to Legal Reasoning.* University of Chicago Press, 1948.

A. R. Lodder. *DiaLaw: on legal justification and dialogical models of argumentation.* Kluwer Academic Publishers, Dordrecht, 1999.

A. R. Lodder and A. Herczog. DiaLaw: A dialogical framework for modeling legal reasoning. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 146–155. ACM Press, 1995.

R. Loui, J. Norman, J. Olson, and A. Merrill. A design for reasoning with policies, precedents, and rationales. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 202–211. ACM Press, 1993.

R. P. Loui and J. Norman. Rationales and argument moves. *Artificial Intelligence and Law*, 3:159–189, 1995.

R. P. Loui, J. Norman, J. Altepeter, D. Pinkard, D. Craven, J. Linsday, and M. Foltz. Progress on Room 5: A testbed for public interactive semi-formal legal argumentation. In *Proceedings of the 6th International Conference on Artificial Intelligence and Law*, pages 207–214. ACM Press, 1997.

L. S. Lutomski. The design of an attorney's statistical consultant. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, pages 224–233. ACM Press, 1989.

D. N. MacCormick and R. S. Summers. *Interpreting Statutes: A Comparative Study.* Routledge, 2016.

J. D. Mackenzie. Question-begging in non-cumulative systems. *Journal of philosophical logic*, pages 117–133, 1979.

J. Maranhão, E. G. de Souza, and G. Sartor. A dynamic model for balancing values. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 89–98. ACM Press, 2021.

C. C. Marshall. Representing the structure of a legal argument. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, pages 121–127. ACM Press, 1989.

L. T. McCarty. Reflections on TAXMAN: An experiment in artificial intelligence and legal reasoning. *Harvard Law Review*, 90:837, 1976.

L. T. McCarty. A language for legal discourse i. basic features. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, pages 180–189. ACM Press, 1989.

L. T. McCarty. An implementation of Eisner v. Macomber. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 276–286. ACM Press, 1995.

L. T. McCarty and N. Sridharan. The representation of an evolving system of legal concepts: II. phototypes and deformations. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 246–253. ACM Press, 1981.

M. Medvedeva, M. Vols, and M. Wieling. Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, pages 1–30, 2019.

M. Minsky. A framework for representing knowledge. In W. P, editor, *The Psychology of Computer Vision*, pages 211–277. McGraw Hill: New York, 1975.

R. Mochales and M.-F. Moens. Argumentation mining. *Artificial Intelligence and Law*, 19:1–22, 2011.

S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5(1):31–62, 2014.

M.-F. Moens, E. Boiy, R. M. Palau, and C. Reed. Automatic detection of arguments in legal texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 225–230. ACM Press, 2007.

M. Možina, J. Žabkar, T. Bench-Capon, and I. Bratko. Argument based machine learning applied to law. *Artificial Intelligence and Law*, 13: 53–73, 2005.

J. Mumford, K. Atkinson, and T. Bench-Capon. Machine learning and legal argument. In *Proceedings of the 21st Workshop on Computational Models of Natural Argument. CEUR Workshop Proceedings*, volume 2937, pages 47–56, 2021.

J. Mumford, K. Atkinson, and T. Bench-Capon. Reasoning with legal cases: A hybrid ADF-ML approach. In E. Francesconi, G. Borges, and C. Sorge, editors, *Legal Knowledge and Information Systems. JURIX 2022: The Twenty-fifth Annual Conference*, pages 93–102. IOS Press, Amsterdam, 2022.

M. Neil, N. Fenton, D. Lagnado, and R. Gill. Modelling competing legal arguments using Bayesian model comparison and averaging. *Artificial Intelligence and Law*, 27:403–430, 2019.

D. Nute. Defeasible logic. In D. Gabbay, C. Hogger, and J. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 253–395. Clarendon Press, Oxford, 1994.

D. Odekerken and F. Bex. Towards transparent human-in-the-loop classification of fraudulent web shops. *Legal Knowledge and Information Systems: JURIX 2020*, 334:239–242, 2020.

D. Odekerken, A. Borg, and F. Bex. Efficient argument-based inquiry at the dutch police. *on Applications of AI to Forensics 2020 (AI2Forensics 2020)*, page 22, 2020.

D. Odekerken, F. Bex, A. Borg, and B. Testerink. Approximating stability for applied argument-based inquiry. *Intelligent Systems with Applications*, 16:200110, 2022.

D. Odekerken, F. Bex, and H. Prakken. Precedent-based reasoning with incomplete cases. In G. Sileno, J. Spanakis, and G. van Dijck, editors, *Legal Knowledge and Information Systems. JURIX 2023: The Thirty-Sixth Annual Conference*, pages 33–42. IOS Press, Amsterdam, 2023a.

D. Odekerken, F. Bex, and H. Prakken. Justification, stability and relevance for case-based reasoning with incomplete focus cases. In *Proceedings if the 19th International Conference on Artificial Intelligence and Law.* ACM Press, 2023b.

R. M. Palau and M.-F. Moens. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 98–107. ACM Press, 2009.

N. Pennington and R. Hastie. Reasoning in explanation-based decision making. *Cognition*, 49(1-2):123–163, 1993.

C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric. A Treatise on Argumentation.* University of Notre Dame Press, 1969.

C. Perelman, H. J. Berman, C. Perelman, and H. J. Berman. *Justice and Reason.* Springer, 1980.

J. G. Peters, F. J. Bex, and H. Prakken. Model-and data-agnostic justifications with a fortiori case-based argumentation. In *Proceedings if the 19th International Conference on Artificial Intelligence and Law.* ACM Press, 2023.

J. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person.* MIT Press, Cambridge, MA, 1995.

D. Poole. On the comparison of theories: Preferring the most specific explanation. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, volume 85, pages 144–147. ACM Press, 1985.

H. Prakken. A tool in modelling disagreement in law: preferring the most specific argument. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Law*, pages 165–174. ACM Press, 1991.

H. Prakken. A logical framework for modelling legal argument. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 1–9. ACM Press, 1993.

H. Prakken. From logic to dialectics in legal argument. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 165–174. ACM Press, 1995.

H. Prakken. An exercise in formalising teleological case-based reasoning. *Artificial Intelligence and Law*, 10:113, 2002.

H. Prakken. Analysing reasoning about evidence with formal models of argumentation. *Law, Probability and Risk*, 3(1):33–50, 2004.

H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005a.

H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 85–94. ACM Press, 2005b.

H. Prakken. A formal model of adjudication dialogues. *Artificial Intelligence and Law*, 16:305–328, 2008a.

H. Prakken. Formalising ordinary legal disputes: a case study. *Artificial Intelligence and Law*, 16:333–359, 2008b.

H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.

H. Prakken. Modelling accrual of arguments in ASPIC+. In *Proceedings of the 17th International Conference on Artificial Intelligence and Law*, pages 103–112. ACM Press, 2019.

H. Prakken. An argumentation-based analysis of the Simonshaven case. *TopiCS in Cognitive Science*, 12(4):1068–1091, 2020.

H. Prakken. A formal analysis of some factor- and precedent-based accounts of precedential constraint. *Artificial Intelligence and Law*, 29:559–585, 2021.

H. Prakken and R. Ratsma. A top-level model of case-based argumentation for explanation: formalisation and experiments. *Argument and Computation*, 13:159–194, 2022.

H. Prakken and G. Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4:331–368, 1996.

H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7: 25–75, 1997.

H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6:231–287, 1998.

H. Prakken and G. Sartor. Formalising arguments about the burden of persuasion. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 97–106. ACM Press, 2007a.

H. Prakken and G. Sartor. Formalising arguments about the burden of persuasion. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 97–106. ACM Press, 2007b.

H. Prakken and G. Sartor. More on presumptions and burdens of proof. In E. Francesconi, G. Sartor, and D. Tiscornia, editors, *Legal Knowledge and Information Systems. JURIX 2008: The Twentyfirst Annual Conference*, pages 176–185. IOS Press, Amsterdam, 2008.

H. Prakken and G. Sartor. A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, pages 223–253. Ashgate Publishing, Farnham, 2009.

H. Prakken and G. Sartor. On modelling burdens and standards of proof in structured argumentation. In K. Atkinson, editor, *Legal Knowledge and Information Systems. JURIX 2011: The Twenty-fourth Annual Conference*, pages 83–92. IOS Press, Amsterdam, 2011.

H. Prakken and G. Sartor. Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214–245, 2015a.

H. Prakken and G. Sartor. Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214–245, 2015b.

H. Prakken and G. Sartor. A formal framework for combining legal reasoning methods. In *Proceedings of the 19th International Conference on Artificial Intelligence and Law*, pages 22–236. ACM Press, 2023.

H. Prakken, C. Reed, and D. Walton. Dialogues about the burden of proof. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 115–124. ACM Press, 2005.

H. Prakken, A. Wyner, T. Bench-Capon, and K. Atkinson. A formalization of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 25(5):1141–1166, 2015.

H. Prakken, F. J. Bex, and A. R. Mackor. Editors' review and introduction: Models of rational proof in criminal law. *TopiCS in Cognitive Science*, 12(4):1053–1067, 2020.

C. Reed and T. Norman, editors. *Argumentation Machines. New Frontiers in Argument and Computation.* Kluwer Academic Publishers, Dordrecht, 2004.

C. Reed and G. Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979, 2004.

N. Rescher. *Dialectics: a Controversy-oriented Approach to the Theory of Knowledge.* State University of New York Press, Albany, N.Y., 1977.

A. Rigoni. Representing dimensions within the reason model of precedent. *Artificial Intelligence and Law*, 26:1–22, 2018.

E. Rissland and K. Ashley. A case-based system for trade secrets law. In *Proceedings of the First International Conference on Artificial Intelligence and Law*, pages 60–66. ACM Press, 1987a.

E. L. Rissland. Examples in legal reasoning: Legal hypotheticals. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 90–93. ACM Press, 1983.

E. L. Rissland. Dimension-based analysis of hypotheticals from supreme court oral argument. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, pages 111–120. ACM Press, 1989.

E. L. Rissland and K. D. Ashley. A case-based system for trade secrets law. In *Proceedings of the 1st International Conference on Artificial Intelligence and Law*, pages 60–66. ACM Press, 1987b.

E. L. Rissland and K. D. Ashley. A note on dimensions and factors. *Artificial Intelligence and law*, 10(1-3):65–77, 2002.

E. L. Rissland and M. T. Friedman. Detecting change in legal concepts. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 127–136. ACM Press, 1995.

E. L. Rissland, E. M. Valcarce, and K. D. Ashley. Explaining and arguing with examples. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, pages 288–294, 1984.

E. L. Rissland, D. B. Skalak, and M. T. Friedman. Bankxx: Supporting legal arguments through heuristic retrieval. *Artificial Intelligence and Law*, 4(1):1–71, 1996.

E. L. Rissland, D. B. Skalak, and M. T. Friedman. Evaluating a legal argument program: The bankxx experiments. *Artificial Intelligence and Law*, 5(1):1–74, 1997.

E. L. Rissland, K. D. Ashley, and R. P. Loui. AI and Law: A fruitful synergy. *Artificial Intelligence*, 150(1–2):1–15, 2003.

R. Riveret, A. Rotolo, G. Sartor, R. Bram, and H. Prakken. Success chances in argument games: a probabilistic approach to legal disputes. In A. Lodder and L. Mommers, editors, *Legal Knowledge and Information Systems. JURIX 2007: The Twentieth Annual Conference*, pages 99–108. IOS Press, Amsterdam, 2007.

R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: a game-theoretical investigation. In P. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 324–335. IOS Press, Amsterdam, 2008.

B. Roth, R. Riveret, A. Rotolo, and G. Governatori. Strategic argumentation: a game theoretical investigation. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 81–90. ACM Press, 2007.

A. Rotolo and G. Governatori. Changing legal systems: Legal abrogations and annulments in Defeasible Logic. *Logic Journal of IGPL*, 18: 157–94, 2009.

A. Rotolo, G. Governatori, and G. Sartor. Deontic defeasible reasoning in legal interpretation: two options for modelling interpretive arguments. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 99–108. ACM Press, 2015.

G. Sartor. Teleological arguments and theory-based dialectics. *Artificial Intelligence and Law*, 10(1-3):95–112, 2002.

G. Sartor. Doing justice to rights and values: teleological reasoning and proportionality. *Artificial Intelligence and Law*, 18:175–215, 2010.

G. Sartor, M. Rudnianski, A. Rotolo, R. Riveret, and E. Mayor. Why lawyers are nice (or nasty) a game-theoretical argumentation exercise. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 108–117. ACM Press, 2009.

G. Sartor, D. Walton, F. Macagno, and A. Rotolo. Argumentation schemes for statutory interpretation: a logical analysis. In R. Hoekstra, editor, *Legal Knowledge and Information Systems. JURIX 2014: The Twenty-seventh Annual Conference*, pages 11–20, Amsterdam, 2014. IOS Press.

K. Satoh, T. Kogawa, N. Okada, K. Omori, S. Omura, and K. Tsuchiya. On generality of PROLEG knowledge representatio. In *Proceedings of the 6th International Workshop on Juris-informatics (JURISIN 2012)*, pages 115–128, Miyazaki, Japan, 2012.

R. C. Schank and R. P. Abelson. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Psychology Press, 1977.

M. Schraagen and F. Bex. Argumentation-driven information extraction for online crime reports. In *CKIM 2018 International Workshop on Legal Data Analysis and Mining (LeDAM 2018)*, 2018.

M. Schraagen and F. Bex. Extraction of semantic relations in noisy user-generated law enforcement data. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 79–86. IEEE, 2019.

M. Schraagen, M. Brinkhuis, F. Bex, et al. Evaluation of named entity recognition in dutch online criminal complaints. *Computational Linguistics in the Netherlands Journal*, 7:3–16, 2017.

J. R. Searle. *Rationality in action.* MIT press, 2003.

M. Sergot, F. Sadri, R. Kowalski, F. Kriwaczek, P. Hammond, and H. Cory. The British Nationality Act as a logic program. *Communications of the ACM*, 29(5):370–386, 1986.

M. Shanahan. The event calculus explained. In M. Wooldridge and M. Veloso, editors, *Artificial Intelligence Today. Recent Trends and Developments*, volume 1600 of *Lecture Notes in Computer Science*, pages 409–430. Springer, 1999.

D. Skalak and E. Rissland. Arguments and cases: An inevitable intertwining. *Artificial Intelligence and Law*, 1(1):3–44, 1992.

C. Steging, S. Renooij, and B. Verheij. Discovering the rationale of decisions: Towards a method for aligning learning and reasoning. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 235–239. ACM Press, 2021a.

C. Steging, S. Renooij, and B. Verheij. Rationale discovery and explainable AI. In E. Schweighofer, editor, *Legal Knowledge and Information Systems. JURIX 2021: The Thirty-fourth Annual Conference*, pages 225–234. IOS Press, Amsterdam, 2021b.

C. Steging, S. Renooij, B. Verheij, and T. Bench-Capon. Arguments, rules and cases in law: Resources for aligning learning and reasoning in structured domains. *Argument and Computation*, 14(2):1–8, 2023.

G. Tarello. *L'interpretazione della legge.* Guiffrè, 1980.

B. Testerink, D. Odekerken, and F. Bex. AI-assisted message processing for the netherlands national police. In *ICAIL 2019 Workshop on AI and the Administrative State (AIAS 2019)*, 2019.

S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, and B. Verheij. A structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 109–118. ACM Press, 2015.

S. T. Timmer, J. J. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian Networks. *International Journal of Approximate Reasoning*, 80:475–494, 2017.

S. E. Toulmin. *The Uses of Argument.* Cambridge University Press, 1958.

S. W. van den Braak, G. A. Vreeswijk, and H. Prakken. AVERs: An argument visualization tool for representing stories about evidence. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 11–15. ACM Press, 2007.

F. H. van Eemeren and B. Verheij. Argumentation theory in formal and computational perspective. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, pages 3–73. College Publications, London, 2018.

F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. Snoeck Henkemans, B. Verheij, and J. H. M. Wagemans. *Handbook of Argumentation Theory.* Springer, Berlin, 2014.

B. Van Gijzel and H. Prakken. Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation. *Argument and Computation*, 3:21–47, 2012.

L. van Leeuwen and B. Verheij. A comparison of two hybrid methods for analyzing evidential reasoning. In M. Araszkiewicz and V. Rodríguez-Doncel, editors, *Legal Knowledge and Information Systems: JURIX 2019: The Thirty-second Annual Conference*, pages 53–62. IOS Press, Amsterdam, 2019.

W. van Woerkom, D. Grossi, H. Prakken, and B. Verheij. Hierarchical precedential constraint. In *Proceedings of the 19th International Conference on Artificial Intelligence and Law*, pages 333–342. ACM Press, 2023.

B. Verheij. *Rules, reasons, arguments: formal studies of argumentation and defeat.* Doctoral dissertation University of Maastricht, 1996.

B. Verheij. Automated argument assistance for lawyers. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law*, pages 43–52, 1999.

B. Verheij. Artificial argument assistants for defeasible argumentation. *Artificial intelligence*, 150(1-2):291–324, 2003a.

B. Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346, 2003b.

B. Verheij. Dialectical argumentation with argumentation schemes: An approach to legal logic. *Artificial Intelligence and Law*, 11(2):167–195, 2003c.

B. Verheij. Argumentation support software: Boxes-and-arrows and beyond. *Law, Probability and Risk*, 6:187–208, 2007.

B. Verheij. To catch a thief with and without numbers: Arguments, scenarios and probabilities in evidential reasoning. *Law, Probability and Risk*, 13:307–325, 2014.

B. Verheij. Formalizing value-guided argumentation for ethical systems design. *Artificial Intelligence and Law*, 24(4):387–407, 2016a.

B. Verheij. Correct grounded reasoning with presumptive arguments. In L. Michael and A. Kakas, editors, *15th European Conference on Logics in Artificial Intelligence, JELIA 2016. Larnaca, Cyprus, November 9–11, 2016. Proceedings (LNAI 10021)*, pages 481–496. Springer, Berlin, 2016b.

B. Verheij. Formalizing arguments, rules and cases. In *Proceedings of the 16th International Conference on Artificial Intelligence and Law*, pages 199–208. ACM Press, 2017a.

B. Verheij. Proof with and without probabilities: Correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty. *Artificial Intelligence and Law*, 25:127–154, 2017b.

B. Verheij. Artificial intelligence as law. Presidential address to the Seventeenth International Conference on Artificial Intelligence and Law. *Artificial Intelligence and Law*, 28(2):181–206, 2020a.

B. Verheij. Analyzing the Simonshaven case with and without probabilities. *TopiCS in Cognitive Science*, 12(4):1175–1199, 2020b. URL `https://doi.org/10.1111/tops.12436`.

B. Verheij, J. C. Hage, and H. J. van den Herik. An integrated view on rules and principles. *Artificial Intelligence and Law*, 6(1), 1998.

B. Verheij, F. Bex, S. T. Timmer, C. S. Vlek, J.-J. C. Meyer, S. Renooij, and H. Prakken. Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, 15(1):35–70, 2016.

S. Villata, M. Araszkiewicz, K. Ashley, T. Bench-Capon, L. K. Branting, J. G. Conrad, and A. Wyner. Thirty years of artificial intelligence and law: the third decade. *Artificial Intelligence and Law*, 30(4):561–591, 2022.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. Building Bayesian Networks for legal evidence with narratives: a case study evaluation. *Artificial Intelligence and Law*, 22(4):375–421, 2014.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, 24:285–324, 2016.

G. A. W. Vreeswijk. Formalizing Nomic. working on a theory of communication with modifiable rules of procedure. Technical report, CS 95-02, Vakgroep Informatica (FdAW), Rijksuniversiteit Limburg, Maastricht, 1995.

W. A. Wagenaar, P. J. Van Koppen, and H. F. Crombag. *Anchored narratives: The psychology of criminal evidence.* St Martin's Press, 1993.

D. Walton. *Argumentation schemes for presumptive reasoning.* Lawrence Erlbaum Associates, 1996.

D. Walton. *Legal argumentation and evidence*. Penn State Press, 2002.

D. Walton, G. Sartor, and F. Macagno. An argumentation framework for contested cases of statutory interpretation. *Artificial Intelligence and Law*, 24:51–91, 2016.

D. N. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, 2008.

M. Wardeh, T. Bench-Capon, and F. Coenen. Padua: a protocol for argumentation dialogue using association rules. *Artificial Intelligence and Law*, 17:183–215, 2009.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Exploiting causality in constructing Bayesian network graphs from legal arguments. In M. Palmirani, editor, *Legal Knowledge and Information Systems: JURIX 2018: The Thirty-First Annual Conference*, pages 151–160. IOS Press, 2018.

R. Wieten, F. Bex, H. Prakken, and S. Renooij. Supporting discussions about forensic bayesian networks using argumentation. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, pages 143–152. ACM Press, 2019.

A. Wyner and T. Bench-Capon. Argument schemes for legal case-based reasoning. In A. Lodder and L. Mommers, editors, *Legal Knowledge and Information Systems. JURIX 2007: The Twentieth Annual Conference*, pages 139–149. IOS Press, Amsterdam, 2007.

T. Yuan, D. Moore, and A. Grierson. A conversational agent system as a test-bed to study the philosophical model "DC". In *Proceedings of the 3rd Workshop on Computational Models of Natural Argument (CMNA'03)*, 2003.

J. Zeleznikow and A. Stranieri. The Split-Up system: integrating neural networks and rule-based reasoning in the legal domain. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 185–194. ACM Press, 1995.

H. Zheng, D. Grossi, and B. Verheij. Hardness of case-based decisions: A formal theory. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*, pages 149–158. ACM Press, 2021.