22. Argumentation in AI and law

Giovanni Sartor

1. INTRODUCTION

This chapter introduces AI and law approaches to legal argumentation, showing how such approaches provide formal models that succeed in capturing key aspects of legal reasoning. The chapter is organised as follows. Sections 2, 3, and 4 introduce the motivations and developments of research on argumentation within AI and law. Section 2 looks into the notion of formal inference, and shows how deduction-based approaches fail to account for important aspects of legal reasoning. Section 3 introduces the idea of defeasibility, and argues that an adequate model of legal reasoning should take it into account. Section 4 presents some AI and law models of argumentation. The remaining sections are dedicated to introducing a formal account of argumentation based on AI and law research. Section 5 defines and exemplifies the notion of an argument. Section 6 discusses conflicts between arguments and their representation in argument graphs. Section 7 defines methods for assessing the status of arguments and evaluating their conclusions, and Section 8 summarises the steps from premises to dialectically supported conclusions.

2. FORMAL NON-DEDUCTIVE INFERENCES IN THE LEGAL DOMAIN

The first attempts at formal models of legal argumentation were based on a view of law as a deductive system: a set of premises which (together with the relevant factual premises) would deductively license legal conclusions. Such models supplemented the apparatus of deductive propositional and predicate logic with the tools of modal logics, and of deontic logic (the logic of obligations and permissions) in particular.[1] And deductive logic has been used, more generally, not only to illustrate various issues in legal theory, but also to bring precision into different aspects of legal knowledge and reasoning.[2]

The use of deductive logic in legal contexts, however, has been criticised on different grounds. A first criticism points to the fact that classical logic deals with descriptive sentences—sentences affirming certain facts or combinations of facts—whereas legal norms are prescriptive (or, rather, constitutive: they make it so that certain brute or institutional facts

[1] See Pablo E Navarro and Jorge L Rodríguez, *Deontic Logic and Legal Systems* (Cambridge University Press 2014). See also Chapter 18 by Maribel Narváez Mora in this volume.

[2] As shown by LE Allen, 'Towards a Normalized Language to Clarify the Structure of Legal Discourse' in AA Martino (ed), *Deontic Logic, Computational Linguistics, and Legal Information Systems* (North Holland 1982).

produce further institutional facts, including obligations and permissions).[3] This criticism may also be related to the recent attempt to distinguish causality from mere correlation within science. Leading theorists of causality have argued, following Judea Pearl, that it is a mistake to understand the connection between a cause A and an effect B merely as a correlation consisting in the fact that whenever A happens, B also happens.[4] In fact, in classical logic, the proposition "$A \supset B$" ("if A, then B") is equivalent, by contraposition, to "$\neg B \supset \neg A$" ("if not B, then not A"); yet this equivalence looks absurd if the connective "$\supset$" is understood as expressing a causal relation. Saying, for instance, that a certain drug causes the patient to heal is not equivalent to saying that the fact that the patient does not heal causes her not to take the drug. The same point seems to apply to the connection between operative facts and their normative effects in the law. While we may say that the culpable causation of harm generates, according to the law, an obligation to compensate, it would be nonsensical to say that the absence of an obligation to compensate generates, according to the law, the fact that harm was not caused. It may thus be argued that we need to express the connection between operative facts and legal effect through a different, unidirectional connective that does not allow for contraposition.

But the second, and most important, criticism of the use of deductive logic in legal contexts points to the indefeasibility of deduction. A deductive approach to legal argumentation is appropriate only to the extent that legal premises support certain conclusions indefeasibly or necessarily. As several authors have argued, however, many important argumentative moves in the legal context establish their conclusions only presumptively or defeasibly. Deduction fails to account for some key aspects of legal argumentation; it cannot deal with conflicting points of view or *topoi*, with the interplay of rules and exceptions, or with the dialectics between general principles and specific rules.[5]

Because of the longstanding assumption that *formal* approaches to legal argumentation must necessarily be *deductive*, research on legal "logic", broadly understood, has split into two distinct trends. On the one hand, formal-mathematical approaches have been deployed to capture those parts of legal argumentation that *are* amenable to deduction.[6] As to aspects of legal argumentation that are *not* deductive, on the other hand, the strategy has been to develop informal theoretical approaches.[7] Some legal theorists have accordingly distinguished between two dimensions in legal argument, "external" and "internal" justification,

---

[3] See Jørgen Jørgensen, 'Imperatives and logic' (1937) 7 Erkenntnis 288.

[4] Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Basic Books 2018) ch 1.

[5] On defeasible reasoning, see e.g. John L Pollock, 'Defeasible Reasoning' in Jonathan E Adler and Lance J Rips (eds), *Reasoning: Studies of Human Inference and its Foundations* (Cambridge University Press 2008). On defeasible reasoning in the law, see Giovanni Sartor, 'Defeasibility in Law' in Giorgio Bongiovanni, Gerald Postema, Antonino Rotolo, Giovanni Sartor, Chiara Valentini, and Douglas Walton (eds), *Handbook of Legal Reasoning and Argumentation* (Springer 2018); and Chapter 21 by Francesca Poggi in this volume.

[6] See e.g. Carlos E Alchourrón and Eugenio Bulygin, *Normative Systems* (Springer 1971).

[7] See e.g. Chaïm Perelman and Lucie Olbrechts-Tyteca, *Traité de l'Argumentation: La Nouvelle Rhétorique* (Presses Universitaires de France 1958); translated into English by John Wilkinson and Purcell Weaver as *The New Rhetoric: A Treatise on Argumentation* (University of Notre Dame Press 1969). On informal logic and legal argumentation, see Chapter 19 by Katharina Stevens in this volume.

interpreting this distinction in a particular way.[8] "Internal" justification is taken to be formal: it provides a set of premises that together *deductively* justify a legal conclusion. "External" justification is seen as informal, providing interpretive or factual arguments that support the premises to be deductively used in internal justification.[9]

I will argue in this chapter that, thanks to the new formal approaches developed through research on AI and law, this distinction is no longer needed. These new approaches are able to address formally the topical, dialectical, adversarial, and ampliative aspects of legal reasoning. By a *formal* approach, I mean a technique to assess the validity of an inference by considering only the syntax (or the form) of the inference regardless of its substance. Such a technique does not aim to establish the acceptability of the premises of the inference; it aims, rather, to determine whether there is a valid connection between premises and conclusion: whether acceptance of the premises rationally justifies acceptance of the conclusion.

Note that the notion of inferential *validity* is here extended beyond its standard use within truth-functional deductive logic, where it requires that the truth of the premises semantically entails the truth of their consequence, i.e., that in every logically possible situation in which the premises are true the consequent is also true. This extension is achieved through two complementary ideas: the concept of an inference is enlarged both to include non-deductive or defeasible arguments, and to include not just single inferences but also dialectical relations between inferences.

What is retained, however, is the distinction between inferential *validity*—which concerns only the connection between premises and conclusions—and inferential *soundness*—which also concerns the truth or acceptability of the premises: unacceptable premises can be used in valid inferences, and, conversely, acceptable premises can be used in invalid inferences.

3. FROM LEGAL DEFEASIBILITY TO FORMAL LEGAL ARGUMENTATION

The development of AI and law models of legal argument has been stimulated by the awareness that legal reasoning is typically nonmonotonic—an aspect that was first addressed through the development of nonmonotonic logics.[10]

To understand the difference between monotonic and nonmonotonic reasoning, consider what happens when a set of premises S_0 is expanded with a further set of premises S_1, to obtain the enlarged set $S_0 \cup S_1$ (where "$\cup$" denotes set union). If the reasoning is monotonic, all conclusions derivable from S_0 are also derivable from $S_0 \cup S_1$. If the reasoning is nonmonotonic, that is not necessarily the case: some conclusions derivable from S_0 may no longer be derivable from $S_0 \cup S_1$. In that case, we can say that the conclusions from S_0 that can no longer be

[8] On the distinction, see Jerzy Wróblewski, 'Legal Decision and Its Justification' (1971) 14 Logique et Analyse 409; and Jerzy Wróblewski, 'Legal Syllogism and Rationality of Judicial Decisions' (1974) 5 Rechstheorie 33. See also Chapter 26 by David Martínez-Zorrilla in this volume.

[9] This is, for example, the view informing Neil MacCormick, *Legal Reasoning and Legal Theory* (Oxford University Press 1978).

[10] Early contributions on nonmonotonic reasoning can be found in Matthew L Ginsberg (ed), *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann 1987).

derived from $S_0 \cup S_1$ are *defeated* by the additional premises; and nonmonotonic reasoning is correspondingly also called "defeasible" reasoning. For an example of nonmonotonic reasoning in the law, consider justificatory circumstances in criminal liability. If the only information we have is that the elements of an offence—*actus reus* and *mens rea*—are present in a case, we will correctly conclude that the corresponding punishment is to be administered. But this conclusion must be withdrawn if we come to know that the offence was committed under justificatory circumstances such as, for example, self-defence.[11] And there are many aspects of the law—such as the interaction between rules and exceptions, conflicts between norms, presumptions, temporal reasoning, the dynamics of legal systems, and burdens of proof—in which nonmonotonic reasoning would seem to be at play.[12]

In some early AI and law applications, nonmonotonic reasoning was captured using Prolog, a leading language for logic programming, i.e. for creating computer programmes that consist of sentences in logical form and are executed through logical inferences.[13] In Prolog, defeasible rules can be expressed through negation by failure, usually represented as "*not*". The proposition "*not A*" holds where all possible attempts to prove A (using the information in the programme) fail. Thus, by including negated propositions in the antecedent of a Prolog rule, the rule becomes defeasible: its conclusion can be derived only so long as the negated elements cannot be inferred. For instance, consider a rule such as:

Murder ←intentional Killing, not Justification

This can be read as follows: "There is murder if there is intentional killing and it cannot be established that there is justification" (the symbol "←", often written as ":-" in the usual Prolog syntax, stands for "if" and "," for "and"). This rule enables the Prolog system to conclude *Murder* whenever the programme entails *Murder* but does not entail *Justification*. This conclusion will no longer hold if information on the existence of justification is added to the programme.

Defeasible reasoning can be related to argumentation by considering that an inference (an argument) is defeasible to the extent that it can be successfully attacked by further inferences (arguments). In the case of Prolog, the attack takes place by establishing a proposition the negation of which occurs in the antecedent of a rule. For instance, inferences concluding for *Justification* pre-empt the use of the rule for *Murder*, and thus block all inferences containing that rule. Logic programming languages also exist that allow for rules with conflicting

[11] For an early sketch of this idea, see HLA Hart, 'The Ascription of Responsibility and Rights' (1949) 49 Proceedings of the Aristotelian Society 171.

[12] On defeasibility in the law, see Giovanni Sartor, 'Defeasibility in Law' (n 5). On different ways of understanding defeasibility, see Jordi Ferrer Beltran and Giovanni Battista Ratti (eds), *The Logic of Legal Requirements: Essays on Defeasibility* (Oxford University Press 2012) and Bartosz Brożek, 'Law and Defeasibility: A Few Comments on *The Logic of Legal Requirements*' (2014) 23 Revus 165. See also Luís Duarte d'Almeida, *Allowing for Exceptions: A Theory of Defences and Defeasibility in Law* (Oxford University Press 2015).

[13] The seminal paper on logic programming and the law is MJ Sergot, F Sadri, RA Kowalski, F Kriwaczek, P Hammond, and HT Cory, 'The British Nationality Act as a Logic Program' (1986) 29 Communications of the ACM 370.

conclusions and priorities over such rules, such as Defeasible Logic,[14] which has been extensively deployed in the legal domain.[15]

Subsequent research in AI and law has deepened the analysis of the connection between defeasibility and legal argumentation by shifting and indeed reversing the relation between the two. Rather than focusing on defeasibility and regarding argumentation as one way of accounting for defeasibility, most contributions have focused directly on argumentation, viewing defeasibility as one of its side-effects. This shift of focus fits well with the dialectical nature of legal argumentation, manifest not only in legal controversies (with each party giving arguments that support their position and attack the other party's arguments), but also within doctrinal debates (where arguments are proposed in favour of certain legal interpretations and against alternative views).[16]

4. THE DEVELOPMENT OF ARGUMENTATIVE MODELS WITH AI AND LAW

Research on AI and law has delivered a rich set of approaches to argumentation. Among the earliest attempts to build computational models of arguments, we can mention the TAXMAN system, which partially implemented a model for constructing legal theories and concepts by modifying prototypical exemplars.[17] An early system for reasoning with legal rules was also proposed which applied rules to reach conclusions, but also identified cases of conflicts between competing rules, giving priority to certain rules (those derived from binding case law) over others (those resulting from legal doctrine or restatements).[18]

A number of legal expert systems—system that include a knowledge base of legal rules and concepts, and an inferential engine—have been developed that apply rules deductively,[19] such

[14] Donald Nute, 'Defeasible Logic' in Dov M Gabbai, CJ Hogger, JA Robinson, and D Nute (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol III (Oxford University Press 1994).

[15] For a recent account, see Guido Governatori, Antonino Rotolo, and Giovanni Sartor, 'Logic and the Law' in Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre (eds), *Handbook of Deontic Logic and Normative Systems*, vol 2 (College 2021).

[16] For examples of this transition, see Thomas F Gordon, 'The Importance of Nonmonotonicity for Legal Reasoning' in Herbert Fiedler, Fritjof Haft, and Roland Traunmüller (eds), *Expert Systems in Law: Impacts on Legal Theory and Computer Law* (Attempto 1988), and Thomas F Gordon, 'The Pleadings Game: An Exercise in Computational dialectics' (1994) 2 Artificial Intelligence and Law 239–92. See also Giovanni Sartor, 'The Structure of Norm Conditions and Nonmonotonic Reasoning in Law' in *ICAIL '91: Proceedings of the 3rd International Conference on Artificial Intelligence and Law* (ACM 1991); and Henry Prakken and Giovanni Sartor, 'A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning' (1996) 4 Artificial Intelligence and Law 331.

[17] L Thorne McCarty, 'A computational theory of *Eisner v Macomber*' in Constantino Ciampi (ed), *Artificial Intelligence and Legal Information Systems* (North Holland 1982).

[18] Anne von der Lieth Gardner, *An Artificial Intelligence Approach to Legal Reasoning* (MIT Press 1987).

[19] For an early discussion, see Richard Susskind, *Expert Systems in Law: A Jurisprudential Inquiry* (Oxford University Press 1987).

as the SoftLaw system,[20] the remote ancestor of Oracle Intelligent Advisor, a development environment for expert systems that is commercially distributed. In such systems, the inference (and explanation) of a legal conclusion takes the form of a single argument constructed by chaining the rules whose application leads to the conclusion. For instance, assume that we are given facts A, B and F, and rules (1)$A \wedge B \rightarrow D$ and (2)$D \wedge F \rightarrow C$. Then conclusion C can be achieved by first applying rule (1) to facts A and B to obtain intermediate conclusion D, and then applying rule (2) to D and fact F to obtain C.

Clearly, legal argumentation cannot be fully captured by models such as this, which fail to account for the interaction of conflicting arguments that express competing reasons and points of view. But this does not mean that such deductive systems cannot be—as indeed they have been—successfully deployed in domains (such as taxation, or social security) in which there is a large set of uncontroversial rules to be applied.

While debates on legal defeasibility initially focused on the need to combine rules and exceptions, they came to assume a more general content, addressing the conflict between two ways of representing legal knowledge and reasoning with it. Supporters of the deductive approach held that formal representations of legal knowledge should stick to deduction: each formalised legal norm should be such that its antecedent strictly entails its consequent, even if this requires departing from natural language and from how lawyers address legal issues. Every time a new exception is introduced or detected, all the affected rules should be revised to include, in their antecedent condition, the negation of the exception: for example, given rule $A \rightarrow C$ and exception $B \rightarrow \neg C$, that rule should be transformed into $A \wedge \neg B \rightarrow C$ so as to obtain a consistent representation.[21]

Supporters of a defeasibility approach, by contrast, argued that formal models of the law should remain as close as possible to the actual language of the law, and that formal reasoning methods should as far as possible track actual legal reasoning and argumentation.[22] It is important to stress the difference between an argumentation *basis* and a deductive axiomatic *base*. While a deductive axiomatic base is consistent and flat, an argumentation basis is conflictive and possibly hierarchical: it includes reasons that can clash with one another, reasons for preferring one reason over others, and reasons for applying or not applying certain reasons given particular conditions— as we will see from Section 5 onwards.

The attempt to capture defeasibility through argumentation—thus understood—led to significant results in AI and law. A dialectical logic was defined in which competing arguments can be developed from a knowledge base of facts and rules, and such arguments can be attacked, and indeed defeated, by arguments to the contrary.[23] In this framework, moreover, whether an argument prevails over other arguments can be determined by further arguments, which can be constructed from the same knowledge base. The model can also include general

20 Surend Dayal and Peter Johnson, 'A Web-Based Revolution in Australian Public Administration?' (2000) 1 Journal of Information Law & Technology <https://warwick.ac.uk/fac/soc/law/elj/jilt/2000_1/austlii/dayal/> accessed 2 December 2024.

21 Ways of revising knowledge bases to preserve consistency through change are addressed by the belief revision approach, originally advanced by Carlos E Alchourrón, Peter Gärdenfors, and David Makinson, 'On the Logic of Theory Change: Partial Meet Contraction and Revision Functions' (1985) 50 Journal of Symbolic Logic 510.

22 See Sartor, 'Defeasibility in Law' (n 5).

23 Prakken and Sartor, 'A Dialectical Model' (n 16).

principles for prioritising rules, according, for instance, to temporal recency, hierarchical position of the enacting authority, or legal significance of protected interests. A general model for defeasible reasoning, also including strict rules and indefeasible axioms, has subsequently been developed in the ASPIC system.[24] A similar model, using a different logical framework, was also implemented in the CARNEADES system.[25] These models have been used to address various aspects of legal reasoning, including case-based reasoning,[26] presumptions and burdens of proof,[27] strategies for presenting arguments,[28] evidential reasoning,[29] reasoning with multiple legal systems to address private-international-law issues,[30] and applying exclusionary reasons.[31]

Further research has shifted the focus from arguments to theories, i.e., alternative sets of generalisations concerning a legal domain, from which competing arguments, supporting different values, can be developed. In this framework, theory choice can be guided by epistemic concerns but also by impacts on legal and ethical values.[32] The relation between

[24] Henry Prakken, 'An Abstract Framework for Argumentation with Structured Arguments' (2010) 1 Argument and Computation 93.

[25] Thomas F Gordon and Douglas Walton, 'The Carneades Argumentation Framework: Using Presumptions and Exceptions to Model Critical Questions' in Paul E Dunne and Trevor Bench-Capon (eds), *Proceedings of the 2006 Conference on Computational Models of Argument: Proceedings of COMMA 2006* (IOS Press 2006).

[26] Henry Prakken and Giovanni Sartor, 'Reasoning with Precedents in a Dialogue Game' in *ICAIL-97* (ACM 1997); Henry Prakken, Adam Wyner, Trevor Bench-Capon, and Katie Atkinson, 'A Formalisation of Argumentation Schemes for Legal Case-Based Reasoning in ASPIC+' (2015) 25 Journal of Logic and Computation 1141.

[27] Henry Prakken and Giovanni Sartor, 'Presumptions and Burdens of Proof' in Tom M van Engers (ed), *Legal Knowledge and Information Systems – JURIX 2006: The Nineteenth Annual Conference* (IOS Press 2006) 21–30; Giovanni Sartor, 'Defeasibility and Burdens of Proof' (2022) 8 Rechtsphilosophie 152.

[28] Bram Roth, Regis Riveret, Antonino Rotolo, and Guido Governatori, 'Strategic Argumentation: A Game Theoretical investigation' in *ICAIL '07: Proceedings of the 11th International Conference on Artificial Intelligence and Law* (ACM 2007).

[29] Floris Bex, Henry Prakken, Chris Reed, and Douglas Walton, 'Towards a Formal Account of Reasoning About Evidence: Argumentation Schemes and Generalisations' (2003) 11 Artificial Intelligence and Law 125; Bart Verheij, Floris Bex, Sjoerd T Timmer, Charlotte S Vlek, John-Jules Ch Meyer, Silja Renooij, and Henry Prakken, 'Arguments, Scenarios and Probabilities: Connections Between Three Normative Frameworks for Evidential Reasoning' (2016) 15 Law, Probability and Risk 35.

[30] Phan Minh Dung and Giovanni Sartor, 'The Modular Logic of Private International Law' (2011) 19 Artificial Intelligence and Law 233.

[31] Rason-based logic adoptsa conception of rules as exclusionary reasons along the lines defended by Joseph Raz, 'Reasons for Action, Decisions and Norms' (1975) 84 Mind 481. The antecedent of a rule "if *A*, then *B*" is taken to work in two ways: *A* both triggers the rule's conclusion B, and excludes the relevance of other reasons for or against *B*. See Jaap C Hage, *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic* (Kluwer 1997).

[32] Trevor Bench-Capon and Giovanni Sartor, 'A Model of Legal Reasoning with Cases Incorporating Theories and Values' (2003) 150 Artificial Intelligence 97.

argumentation and audiences was addressed in value-based argumentation, a model in which the values endorsed by the audience enable the solution of conflicts between arguments.[33]

An important line of research, inspired by previous work in informal logic, concerns the identification of reasoning schemes, namely, general patterns that justify defeasible inferences in domains such as evidence or legal interpretation.[34] Consider, for instance, the scheme according to which if an expert says something on a topic pertaining to their expertise, this pronouncement should normally be accepted as true; or the scheme according to which if an interpretation of a statutory provision fits with ordinary language, it should normally be adopted. The endorsement of such inference schemes may be justified by rationality, but also by the fact that they happen to be practised in the relevant community.[35]

While the approaches just described focus on the arguments that can be constructed with a given argumentation basis, AI and law research has also pursued the analysis of the ways in which arguments can be advanced within legal dialogues. This requires the definition of protocols determining what moves are allowed at each stage of a dialectical interaction, where every move consists in the production of arguments (and thus in new propositions being endorsed or rejected).[36]

Various projects have addressed case-based argumentation. The most influential AI and law approach to case-based argumentation models the relevant aspects as factors favouring opposite conclusions on the relevant issues. This approach was defined and implemented in both the HYPO project[37] and Cato, a system aimed at training students in legal reasoning.[38] In cases concerning violations of trade secrets, for example, one can distinguish the factors that favour the conclusion desired by the plaintiff (that there has been a violation) and those that favour the conclusion desired by the defendant (that no violation has taken place). Precedent cases can be used to build arguments supporting either conclusion in a new situation. If a precedent with a certain outcome shares with the new situation some factors supporting that outcome, then it can be cited to support that same outcome in the new situation; if the new situation includes factors against that outcome that were not present in the precedent, then a *distinction* can be made. A precedent definitely—*a fortiori*—supports an outcome in a new situation whenever the new situation (1) includes all factors (and possibly additional ones) that

---

33 Katie Atkinson and Trevor Bench-Capon, 'Value-based argumentation' (2021) 8 FLAP 1543.

34 See Henry Prakken, Chris Reed, and Douglas Walton, 'Argumentation Schemes and Generalizations in Reasoning about Evidence' in *ICAIL '03: Proceedings of the 9th International Conference on Artificial Intelligence and Law* (ACM 2003); Douglas Walton, Christopher Reed, and Fabrizio Macagno, *Argumentation Schemes* (Cambridge University Press 2008); Douglas Walton, Fabrizio Macagno, and Giovanni Sartor, *Statutory Interpretation: Pragmatics and Argumentation* (Cambridge University Press 2021).

35 See Douglas Walton and Giovanni Sartor, 'Teleological Justification of Argumentation Schemes' (2013) 27 Argumentation 111.

36 See, among others, Thomas F Gordon, *The Pleadings Game: An Artificial Intelligence Model of Procedural Justice* (Kluwer 1995).

37 EL Rissland and KD Ashley, 'A Case-Based System for Trade Secrets Law' in *ICAIL '87: Proceedings of the 1st International Conference on Artificial Intelligence and Law* (ACM 1987).

38 Kevin D Ashley and Vincent Aleven, 'Towards an Intelligent Tutoring System for Teaching Law Students to Argue with Cases' in *ICAIL '91: Proceedings of the 3rd International Conference on Artificial Intelligence and Law* (ACM 1991).

support that outcome in the precedent and (2) does not include any additional factor against that outcome.

Further research has expanded this model by considering aspects such as intermediate factors (with lower-level factors supporting an intermediate factor, which itself supports higher-level ones), dimensions (factors that can be satisfied to different degrees), the reference to values to determine the importance of factors, and the automatic detection of factors in legal opinions. Logical analyses of case-based reasoning have also been developed that precisely define the extent to which precedents, modelled through competing factors, constrain future decisions (if such decisions are to be consistent with the precedents).[39]

While defeasible legal argumentation and case-based reasoning were initially studied independently, some approaches have sought to bridge the two perspectives. It has been shown, for instance, that warrant-based argument can capture factor-based reasoning.[40] The basic idea is that factors can be modelled through defeasible warrants supporting the factors' outcome: a factor f for outcome o can thus be modelled as a defeasible warrant $f \Rightarrow o$, and so can any conjunction of factors supporting the same outcome. Moreover, each case having outcome o and including a set of factors F_1 supporting o and a set of factors F_2 against o conveys the following message: factors F_1 prevail over factors F_2, as is the case for the rules based on the conjunction of the factors in F_1 and F_2.

Further research has addressed case-based reasoning through formal models of argumentation, in particular by arguing that the decision of a case conveys not only information on the comparison of opposed sets of factors, but also information on the extent to which deciding one way or another would contribute to the values at stake (namely the values which motivate the recognition of certain factors as reasons supporting certain decisions).[41] From a value-based perspective, each decided case conveys not only a message concerning the relative importance of sets of factors, but also a message on the relative importance of impacts on concerned values.[42] Recent work has also focused on multivalued factors, also called "dimensions" (e.g., the time during which one is in a certain job), where the values assigned to the dimension determine the extent to which the dimension's outcome (e.g., entitlement to compensation in case of dismissal) is supported.[43]

---

[39] John Horty and Trevor Bench-Capon, 'A Factor-Based Definition of Precedential Constraint' (2012) 20 Artificial Intelligence and Law 181.

[40] Prakken and Sartor, 'Reasoning with Precedents in a Dialogue Game' (n 26).

[41] Trevor Bench-Capon, Henry Prakken, Adam Wyner, and Katie Atkinson, 'Argument Schemes for Reasoning with Legal Cases Using Values' in *ICAIL '13: Proceedings of the 14th International Conference on Artificial Intelligence and Law* (ACM 2013); Giovanni Sartor, 'The Logic of Proportionality: Reasoning with Non-Numerical Magnitudes' (2013) 14 German Law Journal 1419.

[42] J Maranhão, Edelcio G de Souza, and Giovanni Sartor, 'A Dynamic Model for Balancinreag Values' in *ICAIL '21: Proceedings of the 18th International Conference on Artificial Intelligence and Law* (ACM 2021).

[43] Trevor Bench-Capon and Edwina L Rissland, 'Back to the future: Dimensions Revisited' in Bart Verheij, Arno R Lodder, Ronald P Loui, and Antoinette J Muntjewerff (eds), *Legal Knowledge and Information Systems – JURIX 2001: The Fourteenth Annual Conference* (IOS 2001); John Horty, 'Reasoning with Dimensions and Magnitudes' (2019) 27 Artificial Intelligence and Law 309.

There have also been attempts at using natural-language-processing techniques to extract factors from cases;[44] and at identifying and classifying arguments according to the reasoning schemes they implement.[45]

5. THE NOTION OF AN ARGUMENT

While the previous sections gave an overview of multiple research lines on argumentation within AI and law, I cannot, in a chapter such as this, provide a detailed account of any one of them. But in order to substantiate the main idea from Section 1—that AI and law research provides a formal account of legal argumentation—I will now present what I regard as a general formal account of legal argumentation that reflects important aspects shared by several of the contributions mentioned so far. My account will be divided into two parts. The first deals with the notion of an argument, and the second deals with relations between arguments.

5.1 One-step Arguments

An argument typically consists of a set of argument *steps*, each of which is an inference, i.e. a transition from a set of premises to a conclusion. Each such transition must be justifiable according to a *warrant* (also called an *inference rule*, an *inference policy*, or an *inference scheme*): if the antecedent *A* of a warrant "if *A*, then *B*" holds, then we are authorised to infer (conclusively or presumptively) the warrant's conclusion *B*.[46] I will assume that the warrant itself is a premise of the inference it authorises, and I will use the term "facts" for all other premises.

An important distinction is between *saturated* and *unsaturated* arguments. Unsaturated arguments are enthymemes: they leave certain premises implicit (and the warrant, in particular, is often left implicit). By contrast, in saturated arguments all premises are explicit. Consider, for instance, the unsaturated argument by a police officer who tells a driver that she is subject to a 100 Euros fine since her speed was 80 kph. Compare it to the saturated argument that the driver has to pay the 100 Euros fine given the fact that her speed was over 80 kph, which is more than 20 kph above the 50 kph speed limit, and given also the legal rule that whoever exceeds the 50 kph limit by more than 20 kph is subject to a 100 Euros fine. In the latter case, the legal rule plays the role of a warrant, enabling—warranting—the transition from facts to conclusion.

Here I shall consider saturated arguments only, since formal reasoning—and our ability to check whether an inference is formally correct by checking the syntax (the form) of

[44] Kevin D Ashley and Stephanie Brüninghaus, 'Automatically Classifying Case Texts and Predicting Outcomes' (2009) 17 Artificial Intelligence and Law 125.

[45] Adam Wyner, Rachel Mochales-Palau, Marie-Francine Moens, and David Milward, 'Approaches to Text Mining Arguments from Legal Cases' in Enrico Francesconi, Simonetta Montemagni, Wim Peters, and Daniela Tiscornia (eds) *Semantic Processing of Legal Texts* (Springer 2010).

[46] On the notion of a warrant, see Stephen E Toulmin, *The Uses of Argument* (updated edn, Cambridge University Press 2003) 91.

its premises and conclusion—requires saturation. For present purposes, checking the syntax means first of all controlling whether the facts do fit the warrant that links them to the conclusion.

The language I will use for formalising legal arguments is based on sets of symbols for *predicates*, *constants* (words with uppercase initials, such as *A*, *B*, ..., and numbers) and *variables* (lowercase letters such as $x, y, \ldots$). By combining a predicate with constants or variables, we build *atomic propositions*. We say that an atomic proposition is *grounded* when it does not contain variables. I will take the liberty of representing atomic propositions by including the constants or variables inside the predicate, rather than after it (as in usual logical formalisms). Thus, rather than, say, "drives At Speed Of Kph$\left(Lisa, 80\right)$", I will write: "Lisa drives at the speed of 80 kph".

We also use the propositional connectives "$\neg$" ("not"), "$\wedge$" ("and"), and "$\vee$" ("or") to construct molecular formulas out of atomic ones. For simplicity, we do not include in our language the quantifiers of predicate logic or any functional expressions.

Our language includes two connectives for expressing inference warrants: "$\rightarrow$" for strict warrants, and "$\Rightarrow$" for defeasible warrants. Thus "$A \rightarrow B$" must be read as "if A, then certainly B"; and "$A \Rightarrow B$" must be read as "if A, then presumably B". We give each warrant a name that precedes it. Thus, warrants have one of the following forms: "$w : A \Rightarrow B$" or "$w : A \rightarrow B$". Here "A" is a propositional formula (the *antecedent* of the warrant), "B" is a propositional formula or a variable (the *consequent* of the warrant), and "w" is an atomic proposition that uniquely identifies the warrant (the *name* of the warrant). Our formalism covers both defeasible $\left(w : A \Rightarrow B\right)$ and strict warrants $\left(w : A \rightarrow B\right)$. Thus we can also say that any warrant will be of the general form

$$w : A \begin{pmatrix} \Rightarrow \\ \rightarrow \end{pmatrix} B$$

where the expression "$\begin{pmatrix} \Rightarrow \\ \rightarrow \end{pmatrix}$" stands for the alternative between $\Rightarrow$ and $\rightarrow$, i.e. between defeasible and strict warrants.

Here is an example of a defeasible warrant:

$$\begin{aligned} &SF\left(x, y\right) : x \text{ drives at the speed of kph } y \wedge y \text{ is higher than } 50 \\ &\Rightarrow x \text{ is subject to a fine of 100 Euros} \end{aligned}$$

The warrant *SF* ("Speed Fine") is defeasible since there can be exceptional circumstances (e.g., the driver was carrying a seriously injured person to hospital) in which this warrant may not apply to a driver driving over 50 kph (so that the driver would not be sanctioned).

The name of *general* warrants, such as "SF" is followed by the list of all variables (x, y) included in the warrant. By substituting such variables with constants, as in the following example, we obtain an *instance* of the warrant:

$$\begin{aligned} &SF\left(Lisa, 80\right) : \text{Lisa drives at the speed of 80 kph } \wedge \text{ 80 is higher than 50} \\ &\Rightarrow \text{Lisa is subject to a fine of 100 Euros} \end{aligned}$$

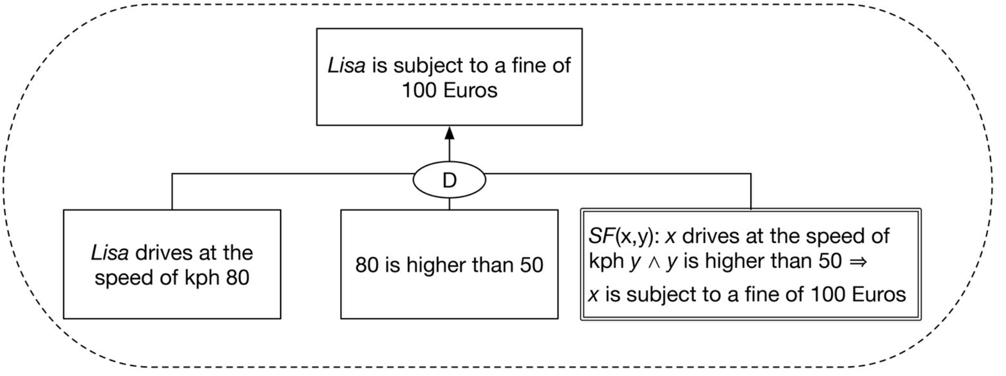

Figure 22.1 An inference graph

By combining a warrant (in this case, *SF*) with factual propositions that match the warrant's antecedent, we obtain a one-step argument, as represented in Figure 22.1, in which the premises are linked though an elliptic connection node to the conclusion they support. Premises and conclusion are included in squares (a double-lined one for the warrant).

We allow for variables within warrants to stand for propositions. This enables us to express reasoning schemes that lead to the endorsement of propositions satisfying certain conditions. Consider, for instance, the (simplified) expert reasoning scheme *EO* ("Expert Opinion") according to which if an expert states a proposition *y*, then *y* is presumably (to be considered as) true; or the reasoning scheme *LA* ("Legislative Authority") according to which if a legislative authority enacts a norm *y*, then *y* holds (as a valid legal content). These two schemes can be represented as:

$$EO(x,y): \quad \text{expert } x \text{ states that } y \Rightarrow y$$
$$LA(x,y): \quad \text{legislative authority } x \text{ enacts norm } y \Rightarrow y$$

5.2 From One-step Arguments to the General Notion of an Argument

We are now well placed to give a general formal notion of an argument. Let us start with the most elementary kind of argument, which consists of just bringing forward any single item, be it a fact or a warrant. For instance, the following are 0-step arguments:

$$A_1: \quad [\text{Lisa drives at the speed of kph } 80]$$
$$A_2: \quad [80 \text{ is higher than } 50]$$
$$A_3: \quad [SF(x,y): x \text{ drives at the speed of kph } y \wedge y \text{ is higher than } 50$$
$$\Rightarrow x \text{ is subject to a fine of 100 Euros}]$$

By combining these three 0-step arguments, we can build the one-step argument A_4, which corresponds to the argument represented in Figure 22.1. Since argument A_4 starts with the three (sub-) arguments A_1, A_2, A_3, and continues with the derivation (indicated by the symbol "$\mapsto$") of a further conclusion (that Lisa is subject to the sanction), we can represent it as follows:

$$A_4 = \left[A_1, A_2, A_3 \mapsto \textit{Lisa}\,\text{is subject to a fine of}\,100\,\text{Euros}\right]$$

This leads us to a general notion of an argument constructible from a knowledge base:

Definition 1 (Argument). *Given a knowledge base* KB *of propositions and warrants:*

- *If A is a proposition or a warrant in KB, then A is an argument from KB.*
- *If* $A_1, \ldots, A_n$ *are arguments from KB having conclusions* $C_1, \ldots C_{n-1}$, C_n *where* C_n, is a warrant having instance *w*: $C_1 \wedge \ldots \wedge C_{n-1} \begin{pmatrix} \rightarrow \\ \Rightarrow \end{pmatrix} C$, *then* $[A_1, \ldots, A_n \mapsto C]$ *is an argument from* KB.
- *Nothing else is an argument from* KB.

Let us consider, for instance, the following knowledge base:

$$\begin{aligned} KB = {} & 1.\ \text{Police officer Tom declares that "Lisa drives at the speed of kph 80"} \\ & 2.\ 80\ \text{is higher than}\ 50 \\ & 3.\ PO(x): \text{Police officer}\ x\ \text{declares that "}y\text{"} \Rightarrow y \\ & 4.\ SF(x,y): x\ \text{drives at the speed of kph}\ y \wedge y\ \text{is higher than}\ 50 \\ & \quad \Rightarrow \text{is subject to a fine of 100 Euros} \end{aligned}$$

Note that *KB* includes, besides the warrant *SF*, also warrant *PO* ("Police Officer"), according to which statements of police officers are presumptively (in the absence of information to the contrary) to be accepted as true.

Based on the above definition, we can construct the following arguments:

$$\begin{aligned} B_1 &= \left[\text{Police officer Tom declares that "Lisa drives at the speed of kph 80"}\right] \\ B_2 &= \left[80\ \text{is higher than}\ 50\right] \\ B_3 &= \left[PO(x,y): \text{Police officer}\ x\ \text{declares that "}y\text{"} \Rightarrow y\right] \\ B_4 &= [SF(x,y): x\,\text{drives at the speed of kph} \wedge \text{is higher than}\,50 \\ & \quad \Rightarrow x\ \text{is subject to a fine of 100 Euros}] \\ B_5 &= \left[B_1, B_3 \mapsto \text{Lisa drives at the speed of kph 80}\right] \\ B_6 &= \left[B_5, B_2, B_4 \mapsto \text{Lisa is subject to a fine of 100 Euros}\right] \end{aligned}$$

In Figure 22.2 you can see all such arguments, each illustrated by a pointer to its conclusion.

Our notion of an argument also enables us to specify the notion of a *subargument*:

Definition 2 (Subargument).

- *If* A *is an argument, then* A *is a sub-argument of* A *(it is useful to assume that any argument is a subargument of itself).*
- *If* $A = A_1, \ldots A_n \mapsto C$ *is an argument, then* $A_1, \ldots A_n$ *are subarguments of* A.
- *If* B *is a subargument of* A, *and* B' *is a subargument of* B *then* B' *also is a subargument of* A.

For simplicity, let us stop our analysis of multilevel arguments at this point. As an example, consider Figure 22.3, where Argument B_6 has subarguments B_6, B_5, B_4, B_3, B_2, B_1. Argument B_4 is based on the warrant *LA* ("Legislative Authority"), according to which what is stated in a legislative text holds in legal reasoning.

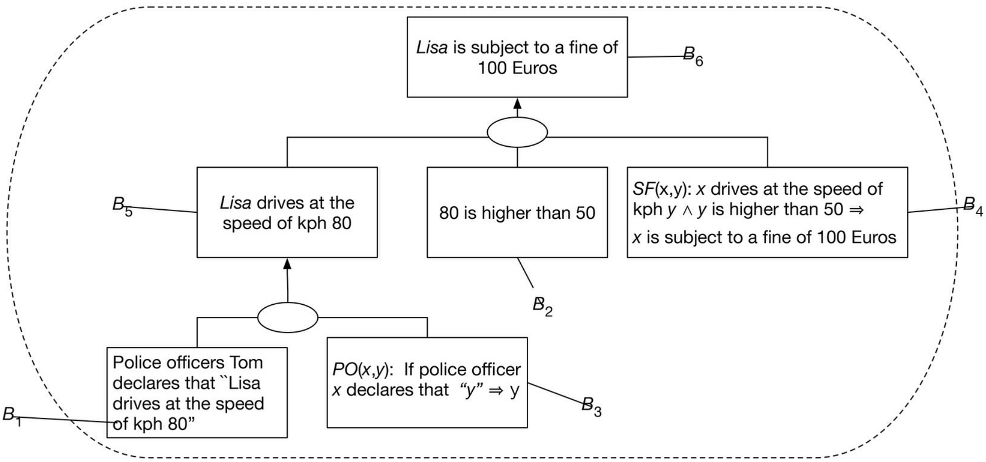

Figure 22.2 A multilevel inference

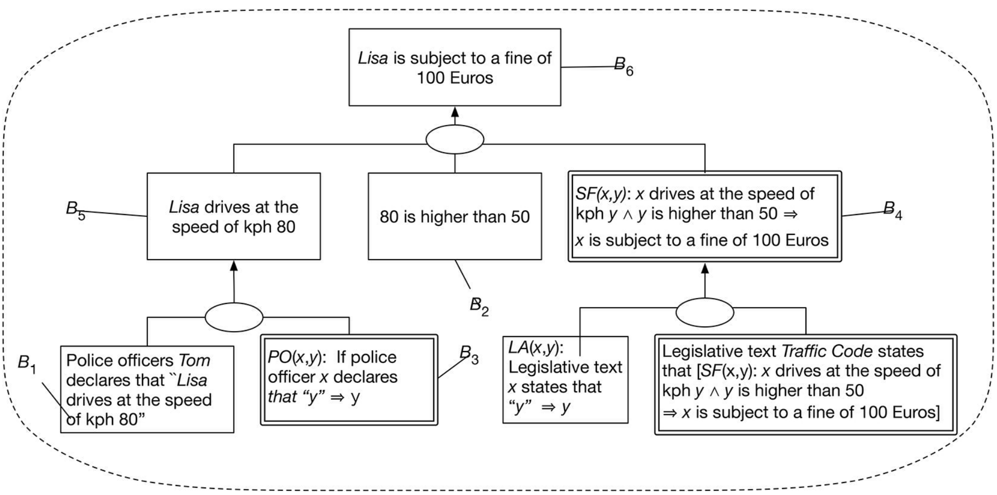

Figure 22.3 An expanded multilevel inference

6. CONFLICTS BETWEEN ARGUMENTS

Let us now consider how arguments can attack other arguments. In the example in Figure 22.4, argument *B* attacks argument *A* by claiming that Lisa is not subject to the fine since her conduct was necessary to prevent a greater harm, based on warrant *Nec* ("Necessity").

This is called a *rebuttal*: the two arguments support incompatible conclusions and therefore attack one another, as represented by the forks in both directions. To solve the conflict, we must determine which argument—which warrant—prevails. In Figure 22.4, it seems that argument *B* should prevail over argument *A*, since warrant *Nec* provides for an *exception* to warrant *SF*.

This is indeed the usual, common-sense interpretation of any statutory text in which a rule with a certain conclusion is followed by further rules specifying that under certain conditions that conclusion will not hold. We find a representation of this in the lower part of Figure 22.5, where argument *C* concludes that argument *B* (on the necessity exception) prevails over argument *A* (establishing the fine), according to warrant *Ex* ("Exception").

By stating this priority (based on warrant *Ex*), argument *C* brings an attack against the attack-link that goes from argument *A* to argument *B*. This reflects the assumption that a weaker argument (*A*), based on an inferior warrant (*SF*), cannot successfully attack a stronger argument (*B*), based on a superior warrant (*Nec*). To compare the strength of competing arguments, we assume the so-called "last-link" principle, according to which the comparative strength of the arguments involved in a rebutting conflict is determined by priority relations over the warrants establishing the incompatible conclusions.[47]

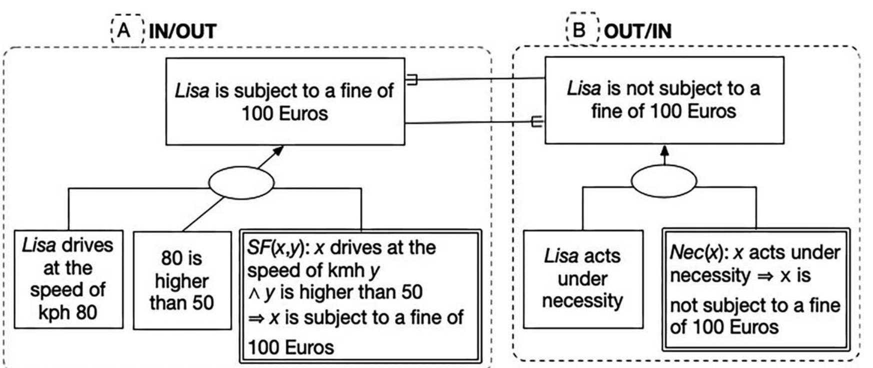

Figure 22.4 A rebutting conflict

[47] On conflicts involving strict rules and on alternative criteria to establish priorities, see Sanjay Modgil and Henry Prakken, 'A General Account of Argumentation with Preferences' (2013) 195 Artificial Intelligence 361.

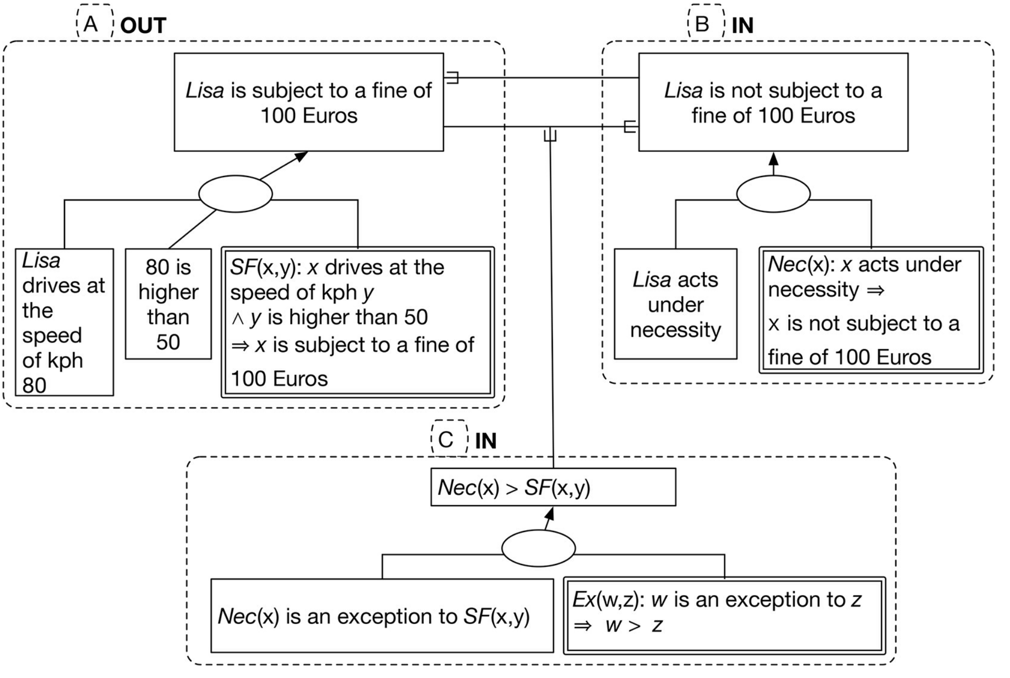

Figure 22.5 A rebutting conflict

Finally, let consider the new attack shown in Figure 22.6.

Here, argument D attacks argument C by claiming that (the warrant *Nec* in) B is not applicable in the given circumstances. This follows, according to warrant *NBM* ("Necessity for Business Meeting"), from the fact that the necessity at stake consists only in the need to arrive in time at a business meeting. Note that the idea that a warrant is not applicable is expressed by denying the name of that warrant: argument D concludes for $\neg Nec\left(Lisa, 100\right)$, i.e., for the inapplicability of warrant *Nec* to Lisa's fine.

Argument D includes an interpretive subargument, which concludes that warrant *NBM* holds (as a legally valid norm), since it contributes to the promotion of legal values. This argument, in turn, is based on the warrant *TI* ("teleological interpretation"), which supports the endorsement of interpretations that contribute to legal values. I do not further develop the interpretive argumentation, which could indeed be extended by specifying what values are promoted by this interpretation, what other canons support different interpretations, and what interpretation is preferable.[48]

Finally, note that an argument A can be attacked also by attacking a subargument of A (remembering that any argument also is a subargument of itself, see Definition 2). Accordingly, we can provide the following definition of an attack.

[48] For an argumentation-based analysis of legal interpretation, see Giovanni Sartor, 'Interpretation, Argumentation, and the Determinacy of Law' (2023) 36 Ratio Juris 214.

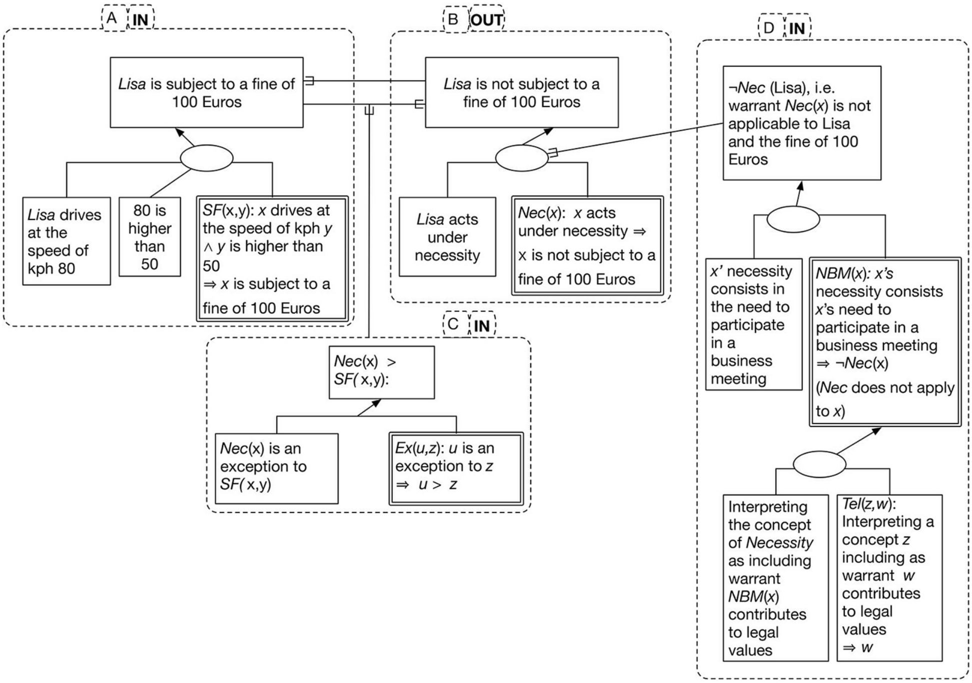

Figure 22.6 An undercutting conflict

Definition 3 (Attack).

- *An argument A attacks an argument B if:*
  - *A rebuts B* (A's conclusion is logically incompatible with the conclusion of a subargument of B); *or*
  - *A undercuts B (A's conclusion affirms that the warrant in a subargument of B is inapplicable).*
- *An argument A attacks an attack-link from argument B to argument C if* A *concludes that B is weaker than C* (B < C).

An *argument graph* is constructed by linking (through arrows) each attacker argument to the argument or link it attacks, as shown in Figures 22.5 and 22.6.

7. ASSESSING THE STATUS OF ARGUMENTS

There are formal approaches to determining the status of arguments relative to an argument graph; i.e., to precisely determine whether an argument is to be accepted, and its conclusion endorsed. Such approaches provide what is usually called "semantics for *abstract* argumentation"; they determine the acceptability of arguments by looking only at attack relations between them, without examining the content of such arguments. Starting with some foundational work carried out in the 1990s,[49] there is now a wealth of research on abstract argumentation, and a range of alternative approaches.[50]

From the perspective of abstract argumentation semantics, the assessment of the merit of arguments—their dialectical evaluation—takes place in the context of the argumentation framework to which they belong: the set of arguments under consideration and their attack relations are modelled through an argumentation graph where attacks are represented through argument links (as in Figures 22.4, 22.5, and 22.6). This buttresses our idea that the logic of arguments is *formal*. It does not address the content of the premises—the propositions and warrants—that are provided as input, but considers only what follows from such premises; what follows, that is, from all the arguments that can be constructed from the premises and the attacks between these arguments, where both the notions of *argument* and *attack* are formally defined.

For assessing arguments, then, let us first define the notion of a *correct status assignment*, which is a formal method for determining what arguments are to be accepted ("IN") or rejected ("OUT") relative to a certain argument graph.

> **Definition 4** (Correct Status Assignment). *A status assignment to the arguments and attack links in an argument graph G is correct if it complies with the following rules:*
>
> - *Any argument A or attack link L in G is IN if no IN argument in G* attacks *A or L through an attack link that is IN.*
> - *Any argument A or attack link L in G is OUT if there is an IN argument in G which attacks A or L through an attack link that is IN.*

By considering all correct assignments of IN/OUT labels to the arguments and attack links in the argument graph, we can evaluate the arguments in the graph:

> **Definition 5** (Argument Evaluation). *The arguments in an argumentation graph can be exhaustively sorted into three disjoint sets:*
>
> - *Arguments that are* ***justified****. These arguments are IN according to all correct assignments of IN/OUT labels to arguments and attacks.*
> - *Arguments that are* ***merely defensible****. These arguments are IN according to some, but not all, correct assignments.*
> - *Arguments that are* ***rejected****. These arguments are OUT according to every correct assignment.*

Note that, directly or indirectly, each *merely defensible* argument must be involved in unresolved conflicts, i.e., conflicts between incompatible arguments that rebut each other, neither of them prevailing over the opponent. In such cases, we may say that, based on the information available in the argument graph, these conflicting arguments are *on a par*: we are unable to determine that either prevails over the other (or that they are equally strong). Without

[49] See Phan Minh Dung, 'On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and n–Person Games' (1995) 77 Artificial Intelligence 321.

[50] See Pietro Baroni, Martin Caminada, and Massimilano Giacomin, 'Abstract Argumentation Frameworks and Their Semantics' in Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre (eds), *Handbook of Formal Argumentation* (College 2018).

information on argument priorities, which would enable us to adjudicate the conflict, we must assume that all outcomes are possible.

Let us now apply the notions specified in Definitions *4* and *5*. Consider Figure 22.4 first. There are no attacks against the attack-links from A to B and from B to A. Therefore, both links are IN. Regarding A and B there are two possibilities. If A is assigned status IN, then B is OUT (being attacked by an IN argument through an IN link). If B is assigned IN, then A is OUT on the same grounds. Thus, we have two correct assignments leading to opposite outcomes regarding the status of A and B. Consequently, according to Definition *4*, both arguments are merely defensible.

Now consider Figure 22.5. We can first observe that argument C, having no attackers, is necessarily IN. Similarly, the attack-link from C toward the attack link from A to B is IN, having no attackers. Then the attack from A against B is OUT. Consequently B is IN, not being attacked through any IN link. This means that A is OUT, being attacked by an IN argument (B) through an IN link. You can check that this is the only correct assignment. This example shows the reinstatement of an OUT argument: argument C, by defeating the attack link from A to B, makes it so that A is IN under all correct assignments, i.e., that A is justified.

Finally, you can also check that the assignment represented in Figure 22.6 is the only one consistent with the rules in Definition *4*: Argument A is justified, since B rejected, given that D is justified.

We now need to determine the status of the claims that can be made relative to an argumentation graph. One idea could be that we should assume that a claim is justified if and only if it is the outcome of a justified argument. Such an approach, however, seems too strict, regarding both factual and interpretive issues, since we need to consider that incompatible defensible arguments may lead to the same conclusion in the matter at stake. In such a case, this conclusion usually may be justifiably endorsed, without the need to make a choice between the competing arguments supporting it. Assume, for instance, that it is uncertain whether Lisa was driving at 80 or at 90 kph when her speed was taken by the police (because, say, two instruments were in use by the police officers that morning, one of which underestimates speed by 10 kph, and it cannot be established which of the two was used on her car). Whether she was driving at 80 or at 90 kph is irrelevant, however, since in either cases there is an argument for fining her.

Definition 6 (The status of argument claims). *Relative to an argument graph* G*:*

- *A claim* ϕ *is **justified** if in every correct IN/OUT assignment to (arguments and attack links in)* G*, there is at least one IN-argument for* ϕ.
- *A claim* ϕi *is **merely defensible** if under some, but not all, correct IN/OUT assignments to* G*, there is at least one IN argument for* ϕ.
- *A claim* ϕ *is **rejected** if under every correct IN/OUT assignment to* G*, all arguments for* ϕ *are OUT.*

According to this definition we may say that, with regard to Figure 22.6, the conclusions are justified that Lisa is subject to the 100 Euros fine and that the *Nec* warrant does not apply to her.

8. FROM ARGUMENT PREMISES TO JUSTIFIED CONCLUSIONS

The framework just presented enables us to specify a logic for formal argumentation, indicating how we may proceed to identify the justified conclusions that follow from a given set of premises.

Consider, for instance, all premises included in the arguments in Figure 22.6. These premises can be identified by the bottom nodes in the graph: all the nodes that are not established through an argument based on further premises (i.e., all the nodes not pointed to by an arrow). The steps that are needed to identify the justified conclusions of such premises are represented in Figure 22.7.

- The first step is to use such premises to construct all possible arguments, according to Definition *1*.
- The second step consists in determining what arguments attack other arguments or attack links, according to Definition *3*, so as to obtain an argument graph.
- The third step consists in evaluating the status of all arguments, based on the status of their attackers, according to Definition *4*.
- The final step consists in evaluating the conclusions of such arguments, according to Definition *5*.

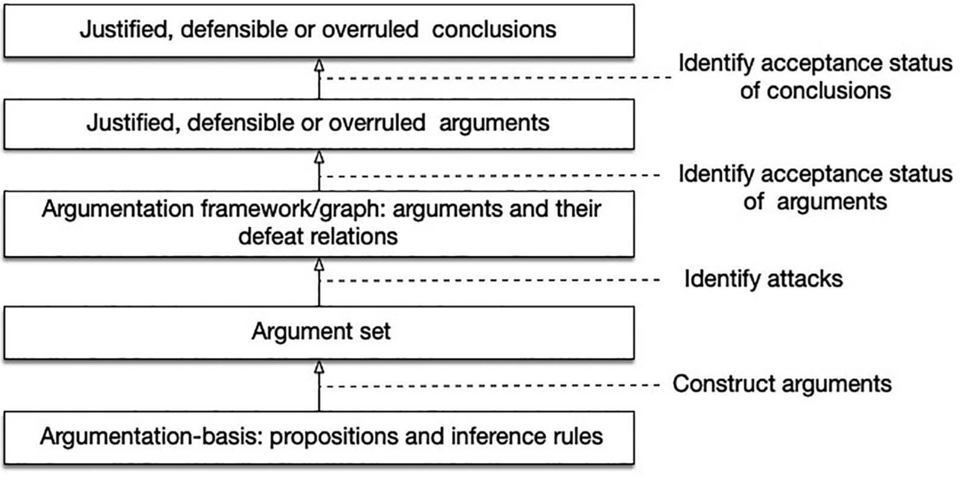

Figure 22.7 The evaluation of argumentation-based claims

Note that this procedure may not be practicable when we have such a large knowledge base that too many arguments would need to be constructed. However, methods can be deployed that focus on the construction of the argument(s) relevant to the issue being addressed.[51]

[51] See for instance, Prakken and Sartor, 'A Dialectical Model' (n 16), where an argument game is provided in which only the arguments relevant to the justification of a certain conclusion are developed.

9. CONCLUSION

In this chapter, besides providing a short account of AI and law research on argumentation, I have discussed what I view as the fundamental achievement of research in this field: the development of formal non-deductive models of legal reasoning. To substantiate the claim that research in AI and law has succeeded in providing formal analyses of non-deductive aspects of legal reasoning, I have provided formal definitions of the following notions: argument, sub-argument, argument attack, evaluation of arguments, and argument-based claims.

The simple formal model presented is sufficiently broad to cover various aspects of legal argumentation addressed by AI and law research (see Section *4*). Some aspects of reasoning with precedents, for example, might be modelled by representing (through warrants) the following ideas: (1) any conjunction of factors leading to the same outcome provides a warrant for that outcome (which will become stronger the larger, according to set inclusion, the set of factors at stake), and (2) any precedent having outcome x and including factors F_1 for x and F_2 against x supports the conclusion that the warrant $F_1 \Rightarrow x$ is stronger than the warrant $F_2 \Rightarrow \neg x$.[52] Similarly, the idea that rules work as exclusionary reasons might be modelled by representing each rule with antecedent A and conclusion B, through two warrants, one stating that A supports conclusion B, and another stating that if A is the case, then the warrants (not qualifying as competing rules) concluding for B or $\neg B$ are inapplicable. I cannot pursue here the idea of using the approach just described as a unifying framework for capturing multiple aspects of legal argumentation, though this seems a promising direction for future research.

While AI and law research on legal argumentation has produced many important theoretical results, it has not yet led to applications having a major impact on legal practice. This is mainly due to the difficulty of formalising legal knowledge (the so-called "knowledge representation bottleneck"), a costly process that often leads to idiosyncratic or anyway dubious results. An opportunity for overcoming this obstacle may lie in the use of natural language processing tools, and in particular of large language models, for the automatic extraction, formalisation and evaluation of legal premises and arguments from legal sources (e.g. from judicial opinions). This could facilitate the construction of transparent argument-based models to support legal analyses, and the construction of legal arguments. Similarly, natural language processing tools can facilitate the deployment of argumentation-based systems by enabling users to interact with such systems in unrestricted natural language. A further opportunity for the use of argumentation-based systems in the legal domain pertains to their integration with machine learning models, to provide explanations of the outcomes of such models, thus contributing to transparency and explainability.[53]

[52] On reasoning with precedent more generally, see also Chapter 5 by Robert Mullins in this volume.

[53] See Kristijonas Čyras, Antonio Rago, Emanuele Albini, Pietro Baroni, and Francesca Toni, 'Argumentative XAI: A Survey' in *Proceedings of the 30th International Joint Conference on Artificial Intelligence* (IJCAI 2021); and Emmanuelle-Anna Dietz, Antonis Kakas, and Loizos Michael, 'Argumentation: A Calculus for Human-Centric AI' (2022) 5 Frontiers in Artificial Intelligence <https://doi.org/10.3389/frai.2022.955579> accessed 2 December 2024.

ACKNOWLEDGEMENT

I am grateful to Ruth Chang and Luís Duarte d'Almeida for their helpful comments on an earlier version of this chapter.