Ratio Juris. Vol. 36 No. 3 September (214–241)

Interpretation, Argumentation, and the Determinacy of Law

GIOVANNI SARTOR*

Abstract. This article models legal interpretation through argumentation and provides a logical analysis of interpretive arguments, their conflicts, and the resulting indeterminacies. Interpretive arguments are modelled as defeasible inferences, which can be challenged and defeated by counterarguments and be reinstated through further arguments. It is shown what claims are possibly (defensibly) or necessarily (justifiably) supported by the arguments constructible from a given interpretive basis, i.e., a set of interpretive canons coupled with reasons for their application. It is finally established under what conditions such arguments provide single outcomes or rather support alternative interpretive conclusions, thus leading to propositions of law whose truth-value is undetermined.

1. Introduction

This article focuses on explicit arguments meant to support or attack a determination of a text's legal content. I will refer to these as "interpretive arguments," without taking up the much-debated distinctions between interpretation strictly understood and other ways of determining legal content—such as construction (Solum 2010), the exercise of judicial discretion (Endicott 2012), or rectification (Soames 2013)—since these ways also involve interpretive arguments broadly understood (see Walton, Macagno, and Sartor 2021, chap. 1).

Legal theorists have proposed different classifications of interpretive arguments. For instance, Tarello (1980) distinguishes the following fourteen kinds of interpretive argument in Italian law: *a contrario*, analogical, a fortiori, from completeness, from coherence, psychological, historical, apagogical, teleological, nonredundancy, authoritative, naturalistic, from equity, from general principles. MacCormick and Summers (1991), in the context of a comparative inquiry, list eleven kinds: from ordinary meaning, from technical meaning, from contextual harmonization, from precedent, from statutory analogy, from a legal concept, from general principle, from history, from purpose, from substantive reasons, from intention. Balkin (2018) lists the following eleven kinds relating to constitutional interpretation in the United States: from text, from structure, from purpose, from consequences, from judicial precedent, from political convention, from the people's customs and lived experience, from natural law or natural rights, from national ethos, from political tradition,

*This article has been supported by the European Research Council (ERC) through the CompuLaw project, carried out under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 833647).

from honoured authority. Here I will not discuss these lists (each of which has the merit of bringing out important patterns of legal reasoning), nor will I attempt an exhaustive classification of arguments, since my focus is rather on providing logical structures that are common to different kinds of interpretive arguments, namely, to interpretive arguments that appeal to different kinds of reasons.[1]

As a running example, I will be using the case of *Dunnachie v Kingston-upon-Hull City Council* (discussed in MacCormick 2005).[2] This case concerns a claim to compensation for moral damages by an employee who had been unfairly dismissed, stating that he had suffered humiliation, injury to feelings, and distress. The key issues to be addressed in this case pertains to the interpretation of Sec. 123(1) of the UK Employment Rights Act 1996, which says that "the amount of the compensatory award shall be such amount as the tribunal considers just and equitable in all the circumstances having regard to the loss sustained by the complainant in consequence of the dismissal." To determine whether the claimant should be compensated not only for his financial losses, but also for his moral harm, it was necessary to determine the scope of the term "loss." On the one hand, the employee argued that an interpretation of this provision in the context of, and in coherence with, all the relevant sections of the statute, should also grant him recovery of *losses* other than financial losses narrowly construed, including moral losses, or injuries to feelings. On the other hand, the employer argued that the relevant section of the UK act only allows for the recovery of *financial loss*, this interpretation corresponding to the ordinary meaning of the term "loss."

2. The Defeasibility of Interpretive Arguments

Interpretive arguments are typically *defeasible*. Their premises only provide *presumptively* sufficient support for their conclusions: If we accept the premises of a valid interpretive argument, we are justified in endorsing the conclusion of that argument, but only so long as the argument is not defeated by relevant counterarguments. Such counterarguments may support conclusions that are incompatible with the conclusion of the argument under attack, or they may exclude the applicably of that argument in the given context.[3]

Consider, for instance, the *Dunnachie* case. On the one hand, a linguistic argument can be constructed for interpreting "loss" in the UK Employment Rights Act as only including pecuniary losses, on the ground that this is the ordinary meaning of the term "loss." On the other hand, a teleological argument could be built for "loss" to also cover moral harm, on the ground that this broader meaning would achieve two goals pursued by the act, namely, better protection for unfairly dismissed workers and a stronger deterrence against unfair dismissals. Both arguments have sound premises, but the conclusions of both cannot be jointly accepted, since they are contradictory: It cannot be the case that "loss" both covers

[1] For an extensive review of interpretive canons in US law, see Scalia and Garner 2012. For an analysis of interpretive argument schemes, see Walton, Macagno, and Sartor 2021.

[2] *Dunnachie v Kingston-upon-Hull City Council* [2004] UKHL 36, [2004] 3 All ER 1011.

[3] Defeasible reasoning is here modelled though argumentation, namely, through the interaction of arguments and counterarguments, an approach that particularly suits legal reasoning (see Sartor 2018). This approach was pioneered by Pollock (1995; 2008) and further developed within AI (for an overview, see Rahwan and Simari 2009) and AI & Law (Gordon and Walton 2009; Prakken and Sartor 2015). For related approaches see also Duarte d'Almeida 2013 and Horty 2012. For a discussion of defeasibility in law see also Brożek 2014 and Ferrer Beltran and Ratti 2012.

and does not cover moral harm. Either we accept one of these conclusions and reject the other, or we remain uncertain and accept neither of them. This shows that interpretive arguments can be defeated: It may be the case that the premises of an interpretive argument are sound, but its conclusion nevertheless cannot be endorsed. This happens when stronger, or at least not weaker, interpretive arguments support incompatible conclusions.

In what follows, I will refer to some interpretive argument schemes—patterns for constructing interpretive arguments—based on different interpretive canons. However, I will first provide a general and most abstract pattern, called *defeasible modus ponens*, under which we can subsume all interpretive schemes.

Defeasible modus ponens presents defeasible arguments in a form that mimics the *modus ponens* scheme of deductive reasoning, namely, the inference pattern: [1] **If** *P*, **then** *Q*. [2] *P*, **therefore** [3] *Q*. Defeasible modus ponens inference has the similar form: [1] **If** *P*, **then presumably** *Q*. [2] *P*, **therefore** [3] *Q*. According to the *presumptive* conditional, *P* only tentatively warrants *Q*: If we accept *P*, then we should also accept *Q*, but only so long as we have no valid reasons to the contrary.

In each of the following examples, the major premise will be a general defeasible conditional which, using the terminology introduced by Toulmin (2003), we may call a *warrant*. The inference consists in matching the warrant's antecedent to the specific facts provided by one or more minor premises, and in deriving the corresponding specification of the consequent of the warrant. For instance, given the warrant (expressed as a defeasible conditional) "If a person is born in Bermuda, then presumably he or she is a British citizen," and the specific fact that "Harry was born in Bermuda," we can defeasibly conclude "Harry is a British citizen" (example from Toulmin 2003, 94). Defeasible inferences may have exceptions and will indeed be defeated if any such exceptions obtain. For instance, the idea that Harry is a British citizen should be abandoned if we come to know that both his parents are citizens of another country.

By modelling different kinds of interpretive and other arguments as defeasible modus ponens inferences, based on different defeasible warrants, we represent all such arguments according to the same abstract logical pattern (see Prakken 2010).

3. The Structure of Interpretive Arguments

In this section, I will present a general structure for interpretive arguments. I will first propose a canonical form for interpretive claims and warrants, and then a corresponding way to support interpretive claims by applying interpretive canons. I will examine how interpretive considerations can be embedded in multistep arguments so as to construct substantive arguments or derive legal interpretations or constructions based on the legislator's intention, *a contrario* arguments, and analogical arguments. Finally, I will address the justification of interpretive canons.

3.1. A Canonical Form for Interpretive Claims and Warrants

The basic form for an interpretive claim is the assertion that a term or phrase in a legal document (hereinafter an "expression") should or should not be interpreted as having a certain content, as in the following example:

- The expression "loss" in document "Employment Rights Act" should / should not be interpreted as meaning "pecuniary loss."

In many cases interpretive claims have a more limited scope, i.e., rather than arguing for or against the ascription of a certain meaning, they only address one aspect of the meaning of an expression, typically arguing that this meaning covers or does not cover a certain content (a certain class of entities):

- The expression *E* in document *D* should be interpreted as *including M*.
- The expression *E* in document *D* should be interpreted as *excluding M*.

For instance, it may be claimed that he expression "loss" in the Employment Rights Act should be interpreted as *including* moral harm or otherwise as *excluding* it.

These variations can be given a general scheme in the following form (see Walton, Sartor, and Macagno 2016):

- The expression *E* in document *D* [should | should not] be interpreted as [meaning | including | excluding] *M*.

where square brackets enclose the possible variants to be considered, separated by a vertical bar (" | ").

In particular, by asserting that the expression *E* means *M*, it is claimed that the scope of *E* and that of *M* (their extensions) always coincide, i.e., that all *M*'s are *E*'s and all *E*'s are *M*'s. By asserting that *E* includes *M*, it is claimed that the scope of *M* is always a subset of the scope of *E* (all *M*'s are *E*'s). By asserting that expression *E* excludes *M*, it is claimed that the scope of *M* is always included in the complement of *E* (no *M*'s are *E*'s).[4] In presenting such patterns, I will henceforth just write "should*" for the alternative "[should | should not]" and "meaning*" for "[meaning | including | excluding]" in the sense just specified.

In legal reasoning, interpretive arguments can take multiple forms according to different interpretive canons. As every canon has the function of either recommending or proscribing certain interpretive choices, we can capture this common function though defeasible conditionals linking the fact that an interpretation fits an interpretive canon to the endorsement of the interpretation recommended by that canon or to the rejection of the interpretation proscribed by it:

(1) **If** interpreting an expression "*E*" in a document "*D*" as meaning* "*M*" fits the canon "*C*," **then presumably** expression "*E*" in document "*D*" should* be interpreted as meaning* "*M*."

For instance, the function of the Ordinary Language canon can be captured by the following defeasible conditional:

[4] Note that the claim that expression *E* excludes *M* is logically stronger than the claim that *E* does not include *M*: The exclusion requires that *all M*'s are not *E*'s, while the noninclusion only requires that *some M*'s are not *E*'s. However, the proposition "*M*'s are not included in *E*'s" would usually be understood as implicating the exclusionary claim that all *M*'s are not *E*'s, rather than the weaker claim that there exists at least one *M* which is not an *E*. For instance, the claim "Merely moral harms are not included in compensable losses" would usually be understood as asserting that no merely moral harm is a compensable loss, rather than as the weaker claim that there exists at least one instance of merely moral harm which is not compensable, while other instances *are* compensable.

(2) **If** interpreting an expression "*E*" in a document "*D*" as meaning* "*M*" fits the canon of Ordinary Language, **then presumably** expression "*E*" in document "*D*" should be interpreted as meaning* "*M*."

3.2. A Canonical Form for Interpretive Arguments

In the format here proposed, arguments link reasons (minor premises) and warrants (major premises) to interpretive conclusions. The reason for adopting or rejecting a proposed interpretation lies in the fact that the interpretation fits a certain canon, the warrant is the canon itself, and the conclusion is the suggestion to endorse or reject the interpretation. We start with a "recommending" example of the argument from Ordinary Language, namely, with the recommendation to adopt a certain interpretation since it fits this canon.

(1) Interpreting the expression "loss" in document "ERA" as meaning "pecuniary deprivation" fits the canon of Ordinary Language. (Minor premise)
(2) **If** interpreting an expression "*E*" in a document "*D*" as meaning "*M*" fits the canon of Ordinary Language, **then presumably** expression "*E*" in document "*D*" should be interpreted as meaning "*M*." (Major premise)

Therefore

(3) The expression "loss" in document "ERA" should be interpreted as meaning "pecuniary deprivation."

Diagram *A* in Figure 1 below represents this interpretive inference, while Diagram *B* in the same figure represents an alternative (and incompatible) interpretive inference supporting the inclusion of injuries to feelings, according to Purposiveness (the ovals contain "P" for "presumably," i.e., they indicate the defeasibility of the inference that leads from the premises below the oval to the conclusion above it).

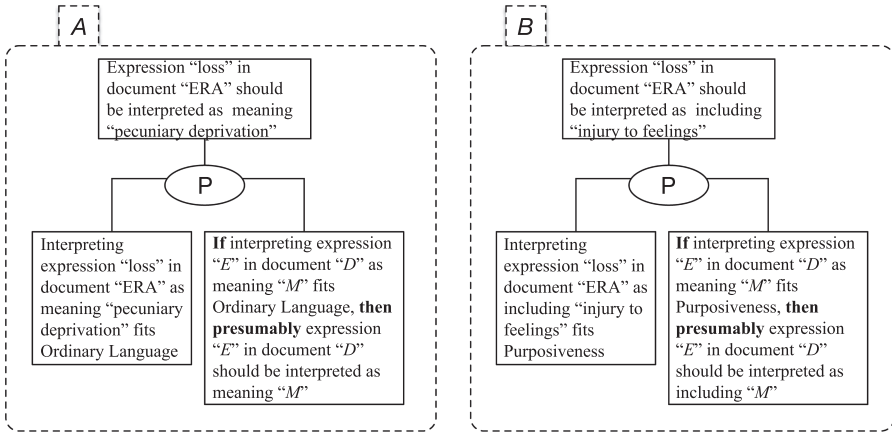

Figure 1. Interpretations according to Ordinary Language and Purposiveness

3.3. Multistep Interpretive Arguments

In the examples above, one-step interpretive arguments have been presented, claiming that an interpretation should be adopted since that interpretation fits a certain canon (Ordinary Language, Purposiveness, etc.). To expand interpretive arguments, we have to combine one-step interpretive arguments with supporting inferences, which provide reasons why that the proposed interpretation does indeed fit the canon.

For instance, the claim that interpreting the expression "loss" as meaning "pecuniary deprivation" (minor premise) fits Ordinary Language, which is a premise in the argument above, can be viewed as the conclusion of the following supporting argument:

(1) The expression "loss" is usually understood by English speakers as meaning "pecuniary loss." (Minor premise)
(2) **If** an expression "*E*" is usually understood by English speakers as meaning "*M*," **then presumably** interpreting "*E*" as meaning "*M*" fits Ordinary Language. (Warrant)

Therefore

(3) Interpreting the expression "loss" in document "ERA" as meaning "pecuniary loss" fits Ordinary Language. (Conclusion)

The two arguments can indeed be chained in the multistep argument *A* shown in Figure 2 below.

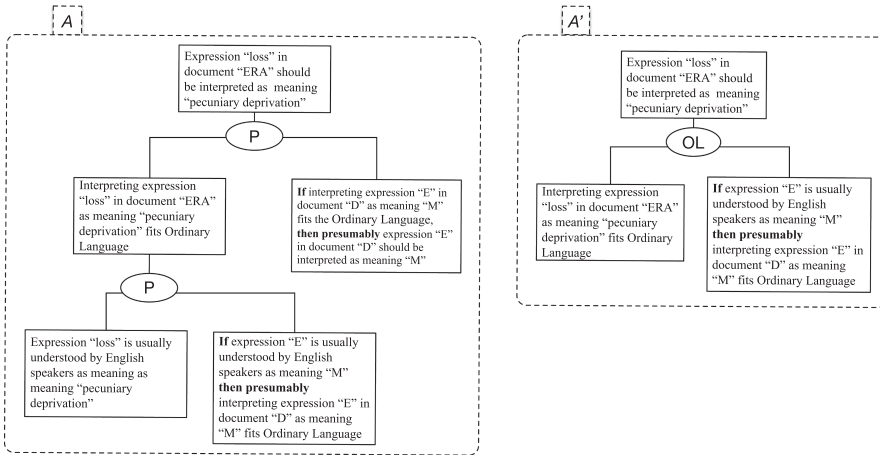

Figure 2. Multistep interpretive argument in extended and compressed form

Note that argument *A*, culminating in an interpretive conclusion, includes a subargument leading to an intermediate conclusion (which works as a premise for the

last inference step in argument *A*). Each of the initial premises of argument *A* (the three bottom leaves of the [inverted] argument tree) can also be viewed as an argument—a most basic kind of argument, one that only consist in asserting a claim without any supporting reasons. It could perhaps be said that there is an implicit reason supporting each such claim, namely, the very fact that the claim was uttered, under the assumption that people usually assert what they believe to be true (and that for the most part they are correct).

Argument *A'* in Figure 2 shows how argument *A* can be compressed (Loui and Norman 1995): A label names the argument scheme being used (OL for Ordinary Language), and the standardised top part in argument *A* is left implicit. Though a compressed form may better suit a synthetic and readable presentation of arguments, here I will use the uncompressed form, which is better suited to my analytical goals.

4. Analysis and Assessment of Conflicting Interpretive Arguments

In this section, conflicts between interpretive arguments will be analysed. This will be done in the context of a general theory of the status of arguments within argumentation frameworks.

4.1. Argument Attacks

We have a conflict between two arguments when at least one of them attacks the other. We can distinguish two kinds of attacks (see Pollock 1995):

- Rebuttal. An argument *A* rebuts an argument *B* if *A*'s conclusion opposes a conclusion in *B*, i.e., if *A*'s conclusion is incompatible with a (final or intermediate) conclusion established by *B*.
- Undercutting. An argument *A* undercuts an argument *B* if *A* opposes an inference in *B*, i.e., if *A* concludes that, under the given circumstances, certain premises in *B* fail to support the conclusion that is linked to such premises.

Figures 3 and 4 below exemplify the two types of attack between competing arguments in the context of legal interpretation:

- Arguments *A* and *B* in Figure 3 rebut each other, since the first concludes that "loss" in "ERA" should be interpreted as meaning "pecuniary detriment" and the second concludes that it should be rather interpreted as to also include "injury to feelings" (the conclusion of the two argument are incompatible).
- Argument *B* in Figure 4 undercuts argument *A*, since *A* reaches a conclusion using teleological reasoning, and *B* argues that teleological inferences do not apply to this case, where there is no ambiguity in Ordinary Language (this argument reflects the disputable textualist assumption that teleological interpretations should be limited to solving ambiguities in ordinary language, an assumption that in its own turn can be attacked, on nontextualist grounds).

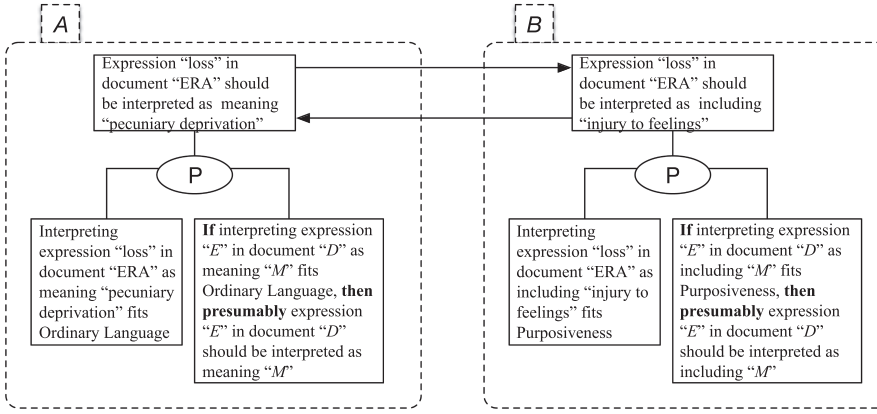

Figure 3. Rebuttal in interpretation

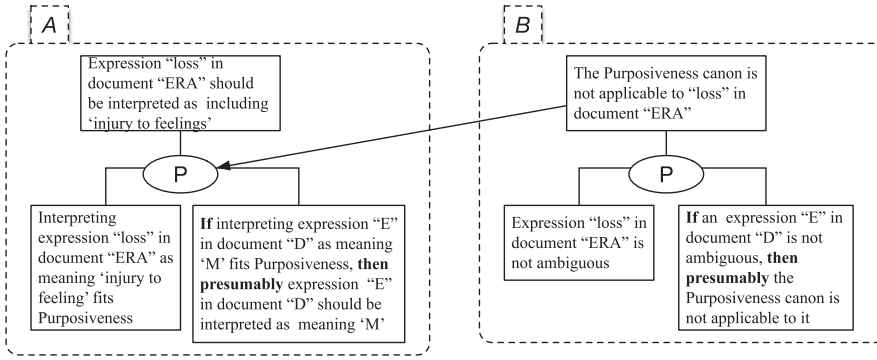

Figure 4. Undercutting in interpretation

Thus, rebuttal concerns arguments *A* and *B* supporting incompatible claims, i.e., it concerns the clash of opposing reasons: *A* presents reasons for φ, while B presents reasons for non-φ (or for a proposition ψ that entails non-φ); undercutting concerns arguments *A* and *B*, such that *B* presents reasons for not taking *A* (its reasons) into account, i.e., for excluding *A*'s relevance to the issue being considered.[5]

Note that, as observed above, even the simple assertion of a claim can be viewed as a limit case of an argument. Thus, according to the foregoing definition, the attack by argument *A* against a premise in an argument *B* (often called "undermining") can also be viewed as case of rebuttal, since the premise in *B* can be viewed as a subargument in it. Consider, for example, the argument that "loss" should be interpreted as "pecuniary loss," since that comports with Ordinary Language, given that this is how "loss" is usually understood by most English speakers. This argument could be undermined by producing some evidence (e.g., a dictionary entry) according to which it is not the case that "loss" is usually understood, by most English speakers, as only including monetary losses; on the contrary, this term is often used to refer to certain nonmonetary detriments (e.g., to permanent bodily harm that does not affect earning capacity).

[5] The idea of undercutting may be linked to the notion of an exclusionary reason as introduced by Raz (1975, sec. 1.2); cf. Nair and Horty (2018, sec. 3.6).

4.2. *From Attack to Reinstatement*

The notion of attack only deals with pairs of arguments; but in argumentation we can have more than two interacting arguments. This leads to so-called reinstatement: An argument *A* that is attacked by an argument *B* may be revived when *B*, in its own turn, is attacked by another argument *C*.

In clarifying this point, it is useful to specify a formal semantics for arguments, spelling out the precise conditions which an argument should meet in order to be acceptable, relative to an argument set. This semantics is based on labelling all arguments in the given argument set with one of two labels: IN (accepted) or OUT (rejected).[6] The basic idea is that an argument is IN if it has no attacker which is IN, while it is OUT if it does have an attacker which is IN. Thus, an attacker which is IN reverts to OUT the argument it attacks, while an attacker which is OUT is not relevant to the state of the arguments it attacks. Thus, we can state the following rules:

- An argument *A* is IN if no argument which attacks *A* is IN.
- An argument *A* is OUT if an argument which attacks *A* is IN.

Let us consider the example in Figure 5 below, which combines rebutting and undercutting attacks. Argument *A* is IN even if attacked by *B*, since *B* is OUT, being attacked by *C*, which is IN, having no attacker. Therefore, having regard to all arguments in

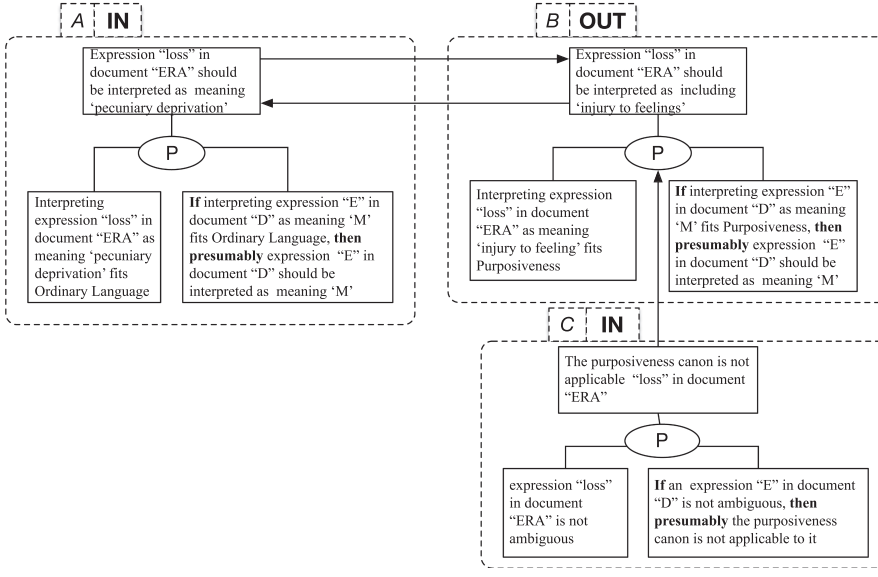

Figure 5. Labelling of arguments, reinstatement

[6] The most influential formal semantics for argumentation has been proposed by Dung (1995) and is based on the idea that an argument *A* in an argument set *S* is acceptable if *A* is included in a subset *E* of *S* that is consistent (no argument in *E* attacks other arguments in *E*) and can respond to all attacks against any arguments in it (if an argument *B* in S attacks any argument in *E*, then there is an argument in *E* that attacks *B*). Here I prefer to use a labelling approach, one that for our purposes is nonetheless equivalent to Dung's semantics (see Baroni, Caminada, and Giacomin 2011).

Figure 5 we should conclude that "loss" should indeed be interpreted as a pecuniary deprivation, since this is the conclusion of the IN-argument *A*.

4.3. Undecided Conflicts, Priorities, and Burdens of Proof

The foregoing analysis of the interaction of arguments needs to be completed with a discussion of the cases in which two arguments rebut each other. Conflicts between two argument contradicting (rebutting) each other can be addressed by considering the comparative strength of the competing arguments: If argument *A* is stronger than (has priority over) the contradictory argument *B*, then *A* should be endorsed notwithstanding the attack, and *B* should be rejected (assuming that no other arguments interfere). This idea may be captured by distinguishing those rebuttals which are effective from those which are ineffective, that is, by assuming that an argument *A* succeeds in rebutting *B* only when *A* is not weaker than (at least as strong as) *B* (as we will see in the next section).

When factual information is at stake, a definite outcome can also be obtained in those cases in which it not possible to determine, with a sufficient degree of certainty, which one of the conflicting arguments is stronger than the other. This result can be reached thanks to burdens and standards of proof. More exactly, if there is a burden of persuasion regarding an operative fact, then the argument for the existence of the fact is relevant only if it is stronger than the competing argument, to the extent that is necessary under the required standard of proof (see Prakken and Sartor 2009; Sartor 2022).

Assume, for instance, that the plaintiff has the burden of proving negligence to establish liability in a tort case: The plaintiff will then lose (the judge will not find the defendant liable) if he can only provide arguments for the defendant's negligence that are not stronger (to the extent required by the applicable standard of proof) than the defendant's arguments against her negligence. Assume, on the contrary, that the defendant has the burden of proving non-negligence: She will lose (the judge will find her liable) if she can only provide arguments against her negligence that are not stronger than the plaintiff's arguments for her negligence.

However, the rules on burden of proof do not usually apply in matters of legal interpretation: When conflicting interpretive arguments lead to opposite conclusions in a legal case, the outcome remains indeterminate, unless the prevalence of one such argument is established. For instance, going back to the interpretation of the expression "loss" in the UK Employment Rights Act, a dazzling palette of alternative interpretations seems possible. The expression "loss" in that act could be interpreted as

- not including injury to feelings according to Ordinary Language;
- including injury to feelings, since it would otherwise be redundant (Nonredundancy);
- not including injury to feelings, to discourage litigation (Purposiveness);
- including injury to feelings, to discourage unfair dismissal (Purposiveness);
- not including injury to feelings, for coherence with other uses of "loss" (Coherence with Language);
- including injury to feelings, for coherence with the constitutional favour for labour (Coherence with Purpose); or
- not including injury to feelings, since this was the legislator's intention (Legislative Intent).

It is easy to see that such arguments support incompatible interpretive claims, i.e., they rebut one another. If no criterion is available for addressing such conflicts (and no other arguments interfere), uncertainty will result: Since each of the fighting argument successfully attacks (defeats) the other, neither can be accepted or rejected. Let us consider the conflict between the Ordinary Language and Purposiveness as represented in Figure 6 below.

Given the information represented in the argument graph in Figure 6, the outcome of the conflict between the two arguments is undetermined: We are unable to establish whether *A* and *B* are IN or OUT. In fact, under the previously outlined labelling rules, there are two possible labelling options. If we assume that *A* is IN, then *B* is OUT, being defeated by an argument which is IN (and *B*, being OUT, is unable to affect the IN state of *A*). If, on the contrary, we assume that *B* is IN, A is OUT. So, there are two possibly correct IN/OUT assignments, and we do not have criteria for choosing between them.

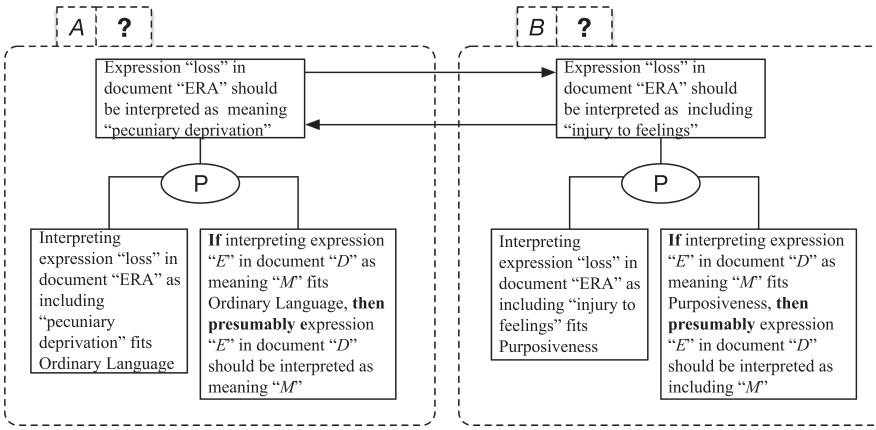

Figure 6. Undecided conflict

4.4. *From Priorities to Priority Arguments*

In the preceding section it was noted that incompatible interpretive arguments may be constructed by using accepted interpretive canons. However, legal cases need to have a decision: Whenever two interpretations are incompatible and choosing one interpretation over the other would entail a different decision, the decision-maker is required to make a choice. Decision-makers may have no explicit reasons for choosing one interpretation over the other: They may choose randomly, or else according to their unarticulated "legal feeling" or "hunch," their "sense of right" (*Rechtsgefühl*). However, this need not always be the case. Since the conflict between two arguments that rebut each other (as in Figure 6) may also be addressed through argumentation, a meta-argument may be provided that

adjudicates the conflict, supporting the choice for one interpretive argument over the other. The nature of this choice may then be determined by the reasons for which preference is accorded. If these reasons are provided by law, we can still say that the law calls for a single interpretive solution. If such reasons are provided by (political) morality—as endorsed by the decision-maker—the interpretive preference will have a moral foundation. If the law itself includes aspects of morality (as argued by "inclusivist" approaches), the preference will be both legal and moral. If the law does not include this aspect (as argued by "exclusivist" approaches), the preference will only be a moral one.

The framework here provided enables us to capture both the case in which the preference between competing arguments is a nonrationalised position of the decision-maker and the case in which it *is* supported by reasons. In the first case the preference would consist in the naked claim that one argument prevails over the other (as we have seen, even a naked claim can be viewed as an argument). In the second case the preference would be the result of a vested argument, where the same claim is supported by reasons. Consider, for instance, the argument set presented in Figure 7. Included in this argument set, apart from the mutually rebutting arguments *A* and *B*, is an additional argument *C* that adjudicates the conflict between *A* and *B*. *C* argues that in this case the Purposiveness argument *B* prevails over the Ordinary Language argument *A*, since the Purposiveness-based interpretation contributes to constitutional values, and the pursuit of constitutional values overrides Ordinary Language considerations.

According to *C*, *B* has the upper hand over *A* (*A* is weaker than *B*). Thus, *B* successfully attacks (i.e., defeats) *A*, while *A*'s attack against *B* is *not* successful. Thus, *C*—the preference argument for *B* over *A*—can also be viewed as an attack against *A*'s attack against *B*. In other words, by claiming that *A* is weaker than *B*, *C* excludes the effectiveness of *A*'s attack against *B* (on preferences as attacks against attacks, see Modgil and Prakken 2013), such that only *B*'s attack against *A* remains, determining an outcome in favour of *B*.

To take into account the possibility that attack links are also attacked, we can refine as follows our rules on IN/OUT assignments:

- An argument *A* or an attack link *L* is IN if no argument which is IN attacks *A* or *L* through an attack link which is IN.
- An argument *A* or an attack link *L* is OUT if an argument which is IN attacks *A* or *L* through an attack link which is IN.

Figure 7 below shows how a priority argument can affect the IN/OUT labels of the arguments whose conflicts it adjudicates. The priority argument *C* states that Purposiveness prevails over Ordinary Language, such that argument *B* prevails over argument *A*. Consequently, *C* successfully attacks the attack-link from the weaker argument (*A*) to the stronger argument (*B*), such that this link is consequently turned into OUT. Since *B* still successfully attacks *A*, while *A* cannot successfully attack *B*, according to our new labelling rules, *B* is now definitely IN, while *A* is definitely OUT.

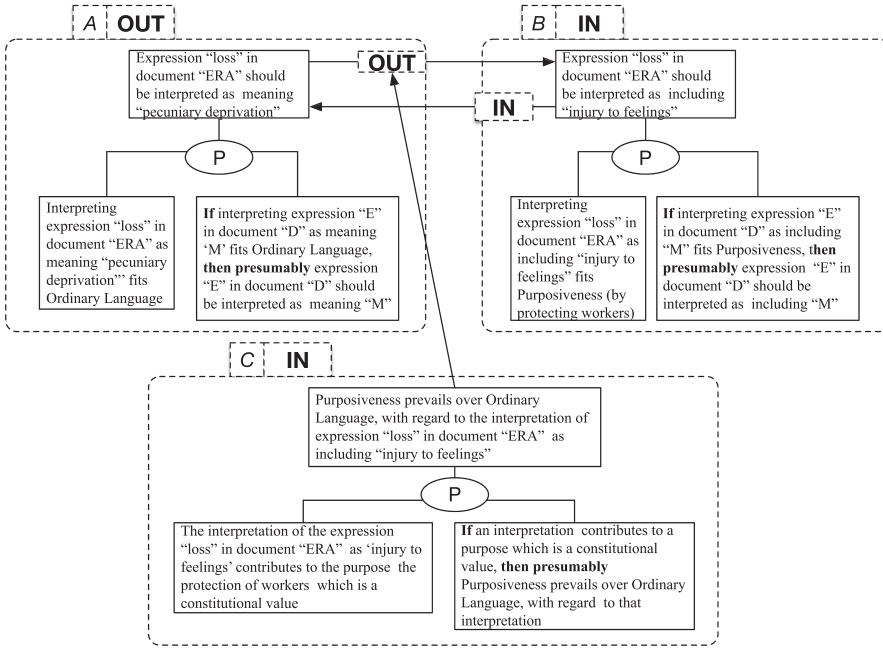

Figure 7. Solution to an interpretive conflict through a priority argument

4.5. *The Accrual of Convergent Arguments*

A multistep argument for an interpretive conclusion may include convergent arguments in favour of applying a single interpretive canon. For instance, consider Justice Scalia's interpretation of the Second Amendment to the US Constitution, i.e., the view that the right to bear arms encompasses the use of weapons for personal defence. A supporter of that view could argue that this interpretation should be adopted on the grounds that it fits with the original meaning of the US constitutional provision being interpreted. She would argue that this is the case since this interpretation corresponds to both (*a*) the ordinary language meaning of that provision at the time it was issued and (*b*) the pragmatics of the act of issuing that provision, given the culture and other clues that were accessible to contemporary readers.

Similarly, an argument for a certain interpretation could also be supported by the convergent application of multiple canons. For instance, the view that "loss" should include injuries to feelings could by supported by the convergence of arguments based on Purposiveness (relative to goals such as the protection of workers) and on Coherence (with other rules allowing for compensation for moral harm).

In both cases, rather than having each argument fight for itself against its counter-argument, arguments having the same conclusion converge and indeed accrue—i.e., reinforce—their joint conclusions: The combination of the separate arguments

supporting the same conclusion often has greater force than that which each such argument would carry separately.[7]

Besides convergences and conflicts between alternative canons, there may also be conflicts and convergences within different applications of the same canon. For instance, different purposes (e.g., legal certainty vs. protecting a weaker party) may support incompatible interpretations according to Purposive interpretations.

The various possibilities are represented in Figure 8 below, which shows the undecided conflict between two interpretive arguments, each resulting from the accrual of convergent interpretive arguments (the formulation of the arguments has been abbreviated to enable them to fit in the space available).

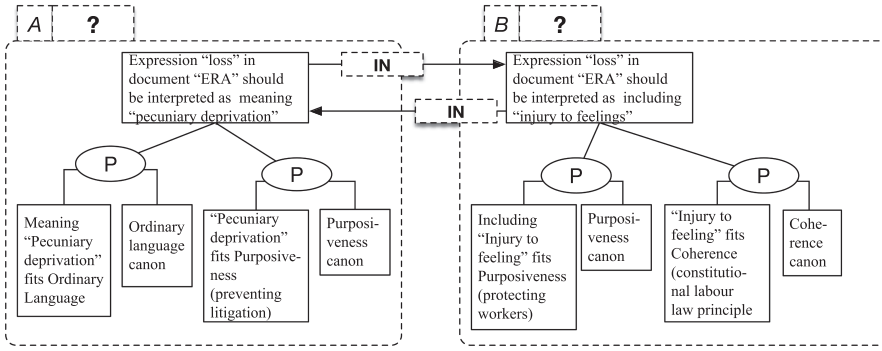

Figure 8. Accrual of arguments

5. Assessing Interpretive Arguments and Conclusions

In this section I will first define the possible dialectical evaluation of arguments and conclusions relative to a given argumentation framework (a set of interacting arguments). I will then analyse the connection among an interpretive basis (a set of interpretive premises), the interpretive arguments that can be constructed from it, and the conclusions supported by such arguments.

5.1. Dialectical Statuses of Interpretive Arguments and Conclusions

The assessment of the merit of arguments—their dialectical evaluation—takes place in the context of the argumentation framework to which they belong, i.e., the set of arguments under consideration and their attack relations. This assessment requires considering all correct assignments of IN/OUT labels to the arguments and attack-links in the framework (by a correct assignment I just mean an assignment consistent with rules 1 and 2 in Section 4.4). On this basis, the arguments in an argumentation framework can be exhaustively sorted into three sets:

- Arguments that are *justified*. These arguments are IN according to all correct assignments of IN/OUT labels to arguments and attacks.

[7] On accrual, see Nair and Horty 2018, sec. 3.5, and for a formal analysis see Prakken 2019.

- Arguments that are *merely defensible*. These arguments are IN according to some, but not all, correct assignments. This is because, whether directly or indirectly, these arguments are involved in an unresolved conflict, i.e., a conflict between incompatible arguments that rebut each other, neither of them prevailing over the opponent, in the context of the given argument set.
- Arguments that are *rejected*. These arguments are OUT according to *any* correct assignment.

On the basis of the assessment of the arguments in an argumentation framework, we can evaluate their conclusions, and more generally the *claims* that can be made, relative to an argumentation framework. As a first approximation, we can consider the following classification:

- A claim is *justified* if it is supported by (i.e., it is the conclusion of) a justified argument.
- A claim is *merely defensible* if it is supported by a merely defensible argument and is not supported by any justified argument.
- A claim is *unfounded* if all arguments supporting it are rejected (this also includes the case in which there is no argument for the claim).

This classification, however, is not fully satisfactory; it requires an adjustment, since a conclusion may be justified (not giving rise to any doubts) not only when it is supported by a justified argument but also when different merely defensible arguments for it exist that cover all relevant alternatives, one such argument being present in each correct IN/OUT assignment. For instance, assume that it is undetermined whether "loss" is to be interpreted as "pecuniary detriment" or as "pecuniary detriment or harm to feelings," such that the arguments for such interpretations are merely defensible (their conflict being undecided). Let us assume that an unfairly dismissed worker only asks to be compensated for the monetary loss he has suffered. It seems that we should conclude that his request is justified, since both dubious (merely defensible) interpretations equally support it: If "loss" means "pecuniary detriment," then it includes monetary losses, and the same holds if "loss" instead means "pecuniary detriment or harm to feelings."

This consideration leads us to the following criteria for evaluating the dialectical status of claim relative to a given argumentation framework:

(1) A claim is *justified* if under every correct IN/OUT assignment there is at least one IN argument supporting that claim (this also includes cases in which different arguments supporting that claim are IN under different IN/OUT assignments).
(2) A claim is *merely defensible* if under some, but not all, correct IN/OUT assignments, there is at least one IN argument supporting that claim.
(3) A claim is *rejected* if under every correct IN/OUT assignment, no IN argument supports the claim (all arguments supporting it are OUT).

5.2. Interpretive Canons (and Relevant Facts) as Argumentation Bases

We have so far considered argumentation frameworks, namely, sets of arguments and their interactions, where these interactions consist in argument conflicts giving

rise to attacks. Let us now look at the premise sets that provide the ingredients for constructing an argumentation framework, i.e., in our case, the set of available interpretive canons and the facts matching these canons.[8]

A set of such premises is not a consistent set of deductive axioms but is rather a repository of materials to be used for building competing arguments and counter-arguments. It is an *argumentation basis*, that is, a set of premises that can be used for constructing an argumentation framework.

Figure 9 below illustrates the process for determining the inferential semantics of an argumentation basis. This process gives us the set of all conclusions that are justified or at least merely defensible relative to the set of premises to be used in constructing arguments. We first construct the argumentation framework resulting from the argumentation basis, that is, we construct all arguments that can be obtained by using the premises in the basis, and we identify all attack relations between such arguments. We then determine what arguments and attacks are IN or OUT (for all or some labelling), and we consequently determine the status of each argument, i.e., whether the argument is justified, merely defensible, or rejected relative to the given argumentation basis. Finally, we identify the status of these arguments' *conclusions*, as noted in Section 4.5.

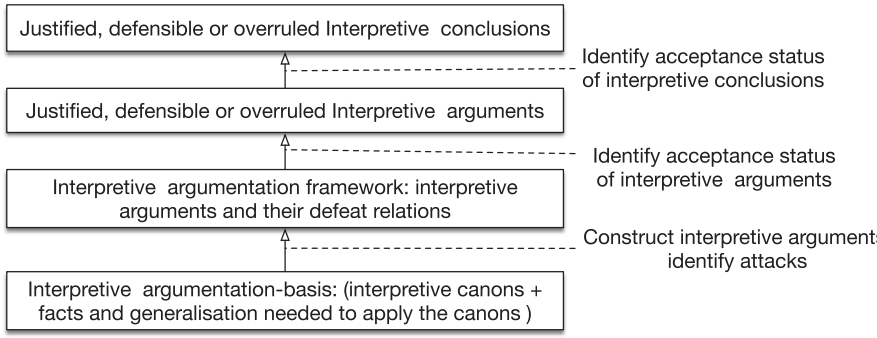

Figure 9. From an interpretive argumentation basis to interpretive conclusions (adapted from Baroni, Caminada, and Giacomin 2011)

5.3. *From Interpretive Argumentation Bases to Interpretive Argumentation Frameworks*

An *interpretive argumentation basis* is the set of relevant interpretive canons, coupled with the textual content of the documents to which the canons are to be applied, along with any further premises that may be relevant to the interpretation of such texts. This argumentation basis determines an *argumentation framework*, which in turn may include multiple interpretive arguments, some of which may be incompatible.

[8] Within formal argumentation, bipolar approaches have also been proposed, which take into account not only attack relations between arguments but also support relations (Amgoud et al. 2008). In the framework here presented support relations are internal to multistep arguments (Section 3.3), each subargument supporting subsequent steps in the argument, or they are addressed through the concept or argument accrual (Section 4.5). The exploration of the possible uses of bipolar approaches in the legal context is a subject for future research.

Let us consider, for instance, the canons of Ordinary Language, Purposiveness, and Coherence. Assume that Ordinary Language favours the interpretation of "loss" as "pecuniary detriment," and so does Purposiveness, in view of the goal of ensuring legal certainty and discouraging litigation. However, in view of the goal of protecting workers, Purposiveness favours, by contrast, the inclusion of "injury to feelings," and so does Coherence, in view of the constitutional principle of the advancement of labour (the Italian Constitution says that the republic ought to promote the empowerment of workers).

Let us further assume that alternative preferences over such arguments are available concerning which of these canons prevails. We then have the situation depicted in Figure 10 below. The IN/OUT status of all arguments in it is indeterminate, i.e., they are all merely defensible. That is because preferences *C* and *D* are incompatible, and there is no criterion for choosing between them (for simplicity's sake, preference arguments have been modelled as unsupported claims, though these claims could as well be supported by reasons). If we assume that *C* is IN, then *D* is OUT, and, consequently, *A* is IN and *B* is OUT. If, on the other hand, we assume that *D* is IN, then *C* is OUT, and, consequently, *A* is OUT, and *B* is IN.

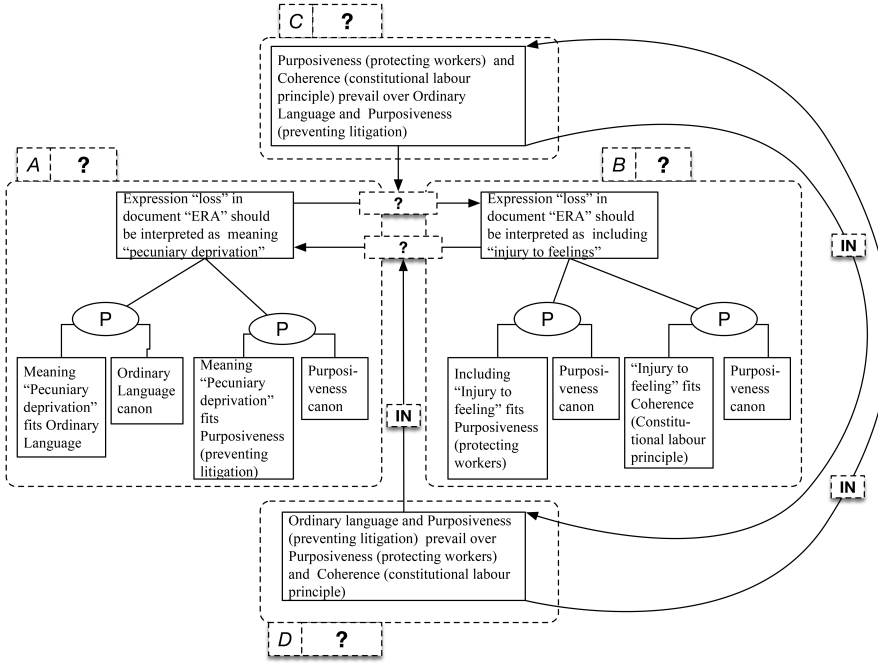

Figure 10. An interpretive framework allowing for alternative, merely defensible solutions

Assume that a legal system only offers the premises for building these interpretive arguments. More specifically, assume that this system does not provide criteria for giving priority to *C* over *D* or vice versa. Under such conditions, we must

conclude that, in the given legal system, arguments *A* and *B* are merely defensible, and so are their incompatible conclusions. We can also describe this situation by saying that both arguments are "legally possible," and that neither is "legally necessary." In fact, the available information does not enable us to decide whether or not to follow the interpretation that (under the canons of Purposiveness and Coherence) favours the substantive constitutional values at stake over the interpretation that (under the canons of Ordinary Language and Purposiveness) favours the "formal" values of certainty and discouragement of litigation.

The situation changes if the argumentation basis is expanded with a metapreference (a preference between preferences) for *C* over *D* as depicted in Figure 11 below. Such a metapreference can be the conclusion of a vested argument; for example, we could argue that interpretations favouring substantive constitutional values prevail since this has been established by a precedent of the constitutional court. The interpretive framework represented Figure 11 has a justified interpretive solution, namely, the conclusion of argument *B*, which is indeed justified in that interpretive framework (while *A* is now rejected). This follows from the fact that according to the newly introduced rightmost argument (which is IN, having no attacker), the preference argument *D* is unable to successfully attack *C*. Thus, *C* is definitely IN, and *A* is consequently considered to be weaker than *B*. Consequently, *A* is OUT, being successfully attacked (defeated) by *B*.

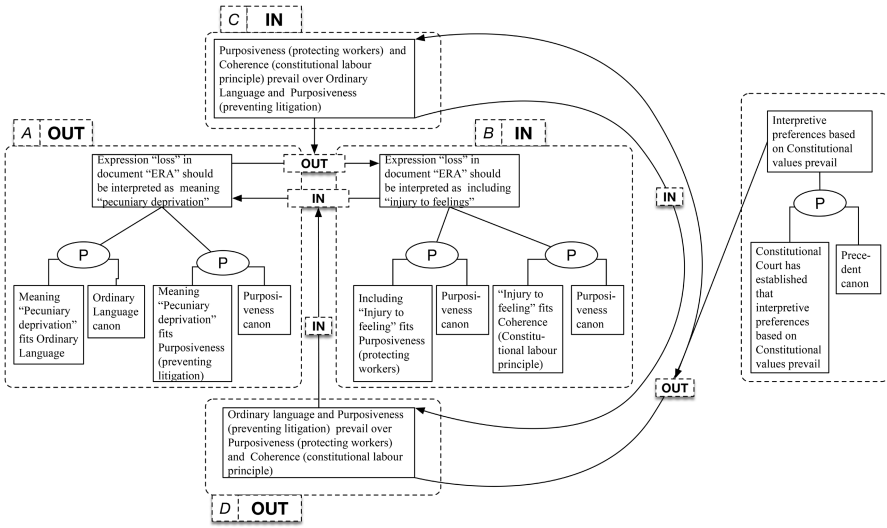

Figure 11. An interpretive framework with a single justified solution

6. Interpretive Arguments and Legal Truths

In this section we will consider to what extent interpretive arguments may determine the truth of propositions of law relative to a given interpretive basis.

6.1. Is There Always One Right Answer to Interpretive Issues?

The foregoing analysis provides a fresh approach to the classic issue of whether and under what conditions there is a determinate answer to an interpretive issue (e.g., the issue of how the term "loss" in the Employment Rights Act should be interpreted).

This issue can have a precise solution when the question concerns a circumscribed, precisely formulated argumentation basis and a corresponding argumentation framework. We can then say that there is a single right answer to an interpretive question, relative to an interpretive basis, if that answer is a justified conclusion of the argumentative framework corresponding to that basis. Conversely, there are *multiple* equally right answers if that framework provides multiple merely defensible answers, namely, alternative defensible conclusions. In the first case, the single justified conclusion can be said to be a *necessary interpretation* relative to the given basis. Necessity consists in the fact that under every correct assignment of IN/OUT labels there is an IN argument for that conclusion. In the second case, each merely defensible conclusion answering the issue may be viewed as a merely *possible interpretation* relative to the given basis. Mere possibility consists in the fact that each alternative answer is supported by IN arguments under *some* assignments but not under all of them. Thus, for instance, relative to the argumentation framework in Figure 10, *alternative* possible (defensible) answers exist concerning the question of whether or not "loss" should be interpreted as including injuries to feelings. On the contrary, relative to the argumentation framework presented in Figure 11, there is a *single* necessary (justified) answer to that question. Note that this analysis of the concept of possible and necessary interpretation corresponds to the standard ideas of necessity as truth in all possible worlds, and possibility as truth in one such world, as the possible worlds we are considering consist in consistent assignments of IN/OUT labels to arguments and attacks constructible from the given interpretive basis.

The information needed to address interpretive issues according to the law includes legal information (interpretive canons, preferences between them, further rules and principles on the application of canons and preferences), as well factual information (including linguistic, social, scientific, and technological knowledge). Whether a legal system delivers a single justified solution or multiple defensible solutions to an interpretive issue depends mainly on two aspects: What positive interpretive criteria—canons, principles, and their priorities—are shared in the legal community, and whether it is not just social sources but also political morality that is viewed as a determinant of the law.

The first aspect is about whether legal criteria exist for deciding conflicts among alternative canons. For instance, a legal system that requires a deferential approach to interpretation—giving preference to Ordinary Language and other linguistic factors, except under exceptional circumstances—would restrict the merely defensible interpretations and expand the justified ones. On the contrary, the range of the merely defensible interpretations would be expanded, and the range of the justified interpretations would correspondingly be restricted, to the extent that priorities among different canons are unavailable or are the object of unresolved conflicts.

The second aspect concerns the extent to which political morality (or general practical reasoning) may contribute to solving interpretive issues (as argued, for

instance, in Alexy 1989 and Dworkin 1986). If political morality could contribute to interpretation, then the range of merely defensible interpretations could in principle be restricted, at least from the standpoint of individual legal reasoners, since their moral arguments may help them solve conflicts between interpretive arguments. Going back to our example concerning the interpretation of "loss," those who believe that the correct moral reading of the Constitution indicates a preference for interpretations favourable to workers, on grounds such as dignity and equality, could solve the indeterminacy in Figure 10 on this basis. However, different people may disagree about the content that should be ascribed to political morality—e.g., some may reject the conception of dignity and equality that leads to that conclusion—or they may disagree about whether substantive principles of political morality should prevail over standards of legal certainty and formal equality. Different moral views may lead their supporters to give different content to the interpretive basis they endorse, and consequently to different views of what the law entails. In other terms, the adoption of different argumentation bases—including different moral principles—would lead different reasoners to construct different interpretive argumentation frameworks supporting different conclusions.

On the contrary, if moral considerations are assumed to be external to the law, all interpretive conflicts that could only be solved based on moral considerations remain unaddressed by the law. From a positive-law perspective, when legal and factual information only supports multiple merely defensible interpretations, the legal analysis should recognise this indeterminacy. This limitation of legal content, however, is compatible with the law allowing, or even mandating, the competent decision-maker to select one of the available merely defensible solutions, drawing on nonlegal considerations—moral, economic, or political.[9] Thus, from this perspective, different reasoners would still come to different determinations in specific cases, but their divergence would be attributed not to different interpretations of the applicable law, but rather to their preference (according to the morality they endorse and consider to be correct or at any rate preferable) about what decisions should be taken under conditions of legal indeterminacy. Under appropriate conditions, such preferences could contribute to the creation of new law, according to the mechanism for the creation of binding precedents, or of customary law.

6.2. Truth-Makers for Interpretive Propositions

The framework here proposed allows for a distinction to be made between *prescriptive interpretive claims* and (in principle) *descriptive interpretive propositions*.[10]

Interpretive claims—here modelled in the form "Expression *E* in document *D* should be interpreted as meaning* *M*"—may be viewed as recommendations directed toward the interpretive community and/or as constitutive declarations (when

[9] This also applies in other circumstances in which the positive law does not provide a definite answer, as argued, among others, by Raz (1979) and Hart (1994).

[10] This distinction may be connected to the well-known distinction between norms and normative propositions proposed by Alchourrón (1969), in the sense that it similarly distinguishes a set of nondescriptive statements (norms) from the assertions made about them. However, this connection does not entail that interpretive claims as well should be viewed as legal norms.

made by the holder of a corresponding power to set interpretations)[11] rather than as propositions that can be true or false depending on how things stand.

Descriptive interpretive propositions, on the other hand, are assertions about interpretive claims, their merit, and their implications. In this section we consider the truth-makers for such propositions, namely, what facts may provide truth conditions for them.

Let us first examine the interpretive statement that, relative to a certain argumentation framework (or basis), a certain interpretive claim is justified, merely defensible, or unfounded. For instance, consider the following proposition: *Relative to the argumentation framework in Figure 10, the claim is justified that the expression "loss" should be interpreted as including "injuries to feelings*."

On our approach, we can say that this proposition is true, since it is indeed the case that its object (the claim whose justification it asserts) is a justified conclusion relative to the argumentation framework in Figure 10. This follows from the argumentation semantics defined in Section 5.1: The proposition that an interpretive claim is (*a*) justified, (*b*) merely defeasible, or (*c*) rejected relative to a certain interpretive framework is true iff that claim is respectively supported by an IN argument under (*a*) all, (*b*) some (but not all), or (*c*) none of the correct IN/OUT assignments under that argumentation framework.

What about the simpler assertion that an expression means*[12] a certain content (or may do so) relative to a determinate interpretive argumentation framework? From the assumption that an expression's legal import can only be determined by the relevant interpretive arguments, it follows that the truth conditions of such an assertion may also fixed by the dialectical status of the corresponding interpretive claims.[13]

Let us first consider the categorical claim that an expression means* a certain content relative to a certain argumentation framework. Such an assertion is true iff the claim is justified that the expression should be interpreted as meaning* that content.[14] For instance, the assertion

1(a) *Relative to the argumentation framework in Figure 10, the expression "loss"* includes *"injuries to feelings."*

[11] See Chiassoni 2016. Such claims can also be considered truth-functional according to a quasi-realist perspective or to approaches that do not refrain from assigning genuine truth-values to normative claims. See, for instance, Dworkin 1996.

[12] Recall from Section 3.1 that "means*" stands for "[means | includes | excludes]."

[13] Also highlighting the dependence of legal truth on arguments is Patterson (1996), who defends it from a realist perspective. For a normative perspective on the truth/validity of legal assertions, see Alexy 1989. Here the truth, indeterminacy, or falsity of interpretive propositions is relativised to interpretive frameworks according to the argumentation logic previously outlined. The logic and semantics here provided make no assumption about what canons should be included, with what priorities, or under what conditions. This depends on the legal system in question and on the legal-philosophical approach adopted (which determines the extent to which moral, scientific, or common-sense knowledge, as well as general practical and epistemic rationality, can contribute to legal interpretation).

[14] A claim is justified if it is supported by IN arguments under every IN/OUT assignment (see Section 5.1).

is true since it is the case that

1(b) *Relative to the argumentation framework in Figure 10, the claim [that the expression "loss" should be interpreted as* including *"injuries to feelings] is justified*.

Similarly, we can say that the more cautious claim that an expression *may* mean* a certain content, relative to a certain argumentation framework, is true iff the claim that the expression should be interpreted as meaning* that content is merely defensible.[15] For instance, the assertion

2(a) *Relative to the argumentation framework in Figure 6, the expression "loss"* may *include "injuries to feelings."*

is true since it the case that

2(b) *Relative to the argumentation framework in Figure 6, the claim [that the expression "loss" should be interpreted as* including *"injuries to feelings"] is* merely *defensible*.

Note that the qualifier "may" is understood as implicating uncertainty (Grice 1989, 23ff.): The assertion that an expression *may* have a certain meaning, in other words, is assumed to be incompatible with the assertion that the same expression (certainly) *does* have that meaning. Thus, an expression *may* have a certain meaning, rather than *categorically* having it, only if arguments exist that support the corresponding interpretive claim, but these arguments are challenged by nonrejected counterarguments. In such a case we can also say that the truth-value of the corresponding categorical statement is undetermined: We cannot establish, relative to that interpretive base, whether or not the expression at issue has that meaning (whether "loss" includes or does not include "injuries to feelings").

Finally, we can say the assertion that an expression means* a certain legal content is *false*, relative to an argumentation basis, iff the corresponding interpretive claim is unfounded.[16] For instance, it can be said to be false that the expression "compensable loss" in the Italian Civil Code also includes people's suffering for the death of their pets (typically in road accidents), as the arguments advanced to this effect go against prevailing arguments based on established precedent in Italian law.

This tripartite semantic analysis of interpretive propositions can be extended to the *implications* of the content ascribed through interpretation; that is, it can be extended to the propositions of law that are *supported* by the meanings ascribed through interpretation. Consider, for instance, the assertions that unfairly dismissed employees under the Employment Rights Act (a) *have* a right to be compensated for pecuniary deprivation, and (b) *may have* a right to be compensated for injuries to feelings. Assertion (a) is grounded in the proposition that "loss" *includes* pecuniary deprivation, while assertion (b) is grounded in the proposition that "loss" *may include* injuries

[15] A claim is merely defensible if it is supported by IN arguments under an IN/OUT assignment, but not under *all* such assignments (see Section 5.1).
[16] All arguments for it are OUT under every IN/OUT assignment.

to feelings. Their truth conditions therefore reflect the truth conditions of these propositions, as described above.

Since each argumentation basis delivers one and only one argumentation framework (the framework containing all and only the arguments constructible from the basis), the truth conditions for interpretive propositions also apply to the argumentation basis: An interpretive proposition is true relative to an argumentation basis if, and only if, it is true relative to the argumentation framework constructible from that basis.

The analysis just presented provides a semantics for interpretive propositions, but two qualifications are needed. First, the assertion that a certain claim is justified, merely defensible, or unfounded relative to an interpretive basis is indeed relative to that interpretive basis, and can be verified only when that basis is unambiguously specified in all its relevant content. When that is not the case, such an assertion presupposes (and implicitly argues for) an interpretive basis that includes premises sufficient to make the assertion true.

For instance, consider the following assertion: *The claim is justified that "loss" in the Employment Rights Act should be interpreted as pecuniary deprivation*. This assertion implicitly assumes that the UK legal system provides the premises (the text, canons, and preferences) needed to build an argument for that claim, and does not contain premises making it possible to build a valid defeater of that argument. In fact, the assertion that an interpretive claim is justified or defensible relative to a certain legal system involves two distinct assumptions: (a) the assumption that the interpretive claim is justified or defensible relative to a certain interpretive basis (as identified or presupposed by the interpreter) and (b) the assumption that this interpretive basis accurately mirrors the relevant content of the legal system being considered. Both assumptions can be contested: The first can be challenged through logical analysis (on the model here presented); the second can be challenged on the basis of factual and legal considerations.

The second qualification concerns the rigidity of our tripartite semantics, which assumes that in each case a clear assessment is made about whether or not an argument is defeated by counterarguments. This rigidity may be overcome by relying on probabilistic argumentation, making it possible to assign probabilities to statuses of arguments, i.e., to conclude, for instance, that an interpretive argument is IN with probability 0.7, but this will be left to future research (see Riveret, Oren, and Sartor 2020).

6.3. *The Dynamics of Legal Interpretation*

The framework just presented can give us some insights into the dynamics of legal interpretation.

We can distinguish two key movements: a reduction in interpretive indeterminacy or an increase in it. The first movement transitions from multiple defensible interpretations toward a single justified interpretation (or at least toward fewer defensible interpretations). The second transitions from a single justified interpretation (or from fewer defensible interpretations) toward multiple (or a larger set of) defensible ones.

Indeterminacy is reduced when—as a consequence of the availability of new interpretive arguments, or of new arguments on priorities over interpretive arguments—some previously defensible arguments become rejected. A key role in this process is played by new judicial decisions, which enable new arguments from precedent, that is, arguments that an interpretation should be adopted since it fits the rationale

of a precedent. Arguments from judicial precedent (especially from judgments by the highest courts) usually prevail over interpretive arguments based on other canons. By imposing the interpretation adopted in precedent as the only justified one, these arguments eliminate the indeterminacy resulting from conflicting defensible interpretations (this applies both to civil law and common law, though with different degrees of stringency). Assume, in our example, that a top court has recently decided a case by arguing that "loss" does not include injury to feelings and has consequently dismissed the corresponding claim for compensation by an unfairly dismissed employee. Based on this precedent (and assuming that no incompatible precedents are available), a decisive argument can be mounted in favour of the interpretation adopted in the precedent, an argument that is likely to prevail over all arguments to the contrary (switching their dialectical status from defensible to rejected), unless grounds exist for rejecting or overruling the precedent.

On the other hand, the indeterminacy of legal interpretations can increase when new interpretive premises become available that provide for challenges to the so-far prevailing interpretive conclusions (switching their dialectical status from justified to merely defensible). This happens in particular when the legal provisions being interpreted remain in force for a long time. Consider, for instance, the rule in Article 2043 of the Italian Civil Code that establishes the obligation to compensate for unjustly caused harm. For decades this rule was assumed to cover only lost profits, but it was later interpreted as also including permanent harm to health not leading to a pecuniary loss, and subsequently as further including some kinds of suffering not leading to permanent physical impairment.

This evolution can be explained by considering that the relevant argumentation basis changed over time, having been expanded with new premises—such as the constitutional recognition of the right to health and the new content and salience of protections relating to individual health and well-being—and these new premises enabled new arguments to be made that challenged the previously adopted interpretations. These arguments generated new indeterminacies, which lingered on until new consistent judicial decisions were adopted that confirmed such interpretive turns. Court rulings that contradict earlier precedents or that introduce distinctions may contribute to expanding indeterminacies, as long as uncertainty remains as to whether the new rulings overrule or limit the earlier ones.

7. Conclusion

In the foregoing, a partly formal account has been provided laying out the logic of statutory interpretation: Arguments have been represented in natural language, but the relation between arguments has been captured through a labelling-based semantics. The inferential status of interpretive arguments and of their conclusions has been examined accordingly, which has led to considerations pertaining to truth conditions for propositions of law and to indeterminacy in the legal domain.

The logical framework here presented contributes to filling a gap in the literature, as few contributions have so far addressed legal interpretation through formal models (see Araszkiewicz 2013; Rotolo, Governatori, and Sartor 2015; Walton, Macagno, and Sartor 2021), while formal argumentation has already been extensively applied to case-based reasoning (Ashley 1990; Horty 2011; Horty and Bench-Capon 2012; Prakken and Sartor 1998).

It is important to note that the semiformal framework here provided covers statutory interpretation only at a very abstract level: The study analyses the dialectical interactions among interpretive arguments, but it does not address the different cognitive mechanisms concerned in applying interpretive argument schemes, which mechanisms involve logical, linguistic, pragmatic, decision-theoretical, and other aspects.[17] Even at this high level of abstraction, however, the model here presented may provide a useful framework for capturing important aspects of interpretive reasoning and developing corresponding jurisprudential analyses.

CIRSFID-AI
University of Bologna
Via Galliera 3
Bologna 40121
Italy
Email: giovanni.sartor@unibo.it

Law Department
European University Institute
Villa Salviati, Via Bolognese 156
Florence 50139
Italy
Email: giovanni.sartor@eui.eu

References

Alchourrón, C. E. 1969. Logic of Norms and Logic of Normative Propositions. *Logique et analyse*, n.s., 12(47): 242–68. https://www.jstor.org/stable/44083577.

Alexy, R. 1989. *A Theory of Legal Argumentation: The Theory of Rational Discourse as Theory of Legal Justification*. Oxford: Oxford University Press.

Amgoud, L., C. Cayrol, M. C. Lagasquie-Schiex, and P. Livet. 2008. On Bipolarity in Argumentation Frameworks. *Intelligent Systems* 23(1): 1062–93. https://doi.org/10.1002/int.20307.

Araszkiewicz, M. 2013. Towards Systematic Research on Statutory Interpretation in AI and Law. In *Legal Knowledge and Information Systems*. JURIX 2013: The Twenty-Sixth Annual Conference. Ed. K. D. Ashley. Frontiers in Artificial Intelligence and Applications, vol. 259, pp. 15–24. Amsterdam: IOS Press. https://doi.org/10.3233/978-1-61499-359-9-15.

Ashley, K. D. 1990. *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. Cambridge, MA: MIT Press.

Baroni, P., M. Caminada, and M. Giacomin. 2011. An Introduction to Argumentation Semantics. *The Knowledge Engineering Review* 26(4): 365–410. https://doi.org/10.1017/S0269888911000166.

[17] Some of these aspects are addressed in Walton, Sartor, and Macagno 2018 and Maranhão and Sartor 2019. See also Brewer 2011; Marmor and Soames 2011; and Walton, Macagno, and Sartor 2021.

Balkin, J. M. 2018. Arguing about the Constitution: The Topics in Constitutional Interpretation. *Constitutional Commentary* 33: 145–255. https://scholarship.law.umn.edu/concomm/1177.

Brewer, S. 2011. Logocratic Method and the Analysis of Arguments in Evidence. *Law, Probability and Risk* 10(3): 175–202. https://doi.org/10.1093/lpr/mgr013.

Brożek, B. 2014. Law and Defeasibility: A Few Comments on "The Logic of Legal Requirements." *Revus* 23: 165–70. https://doi.org/10.4000/revus.3110.

Chiassoni, P. 2016. Legal Interpretation without Truth. *Revus* 29: 93–118. https://doi.org/10.4000/revus.3615.

Duarte d'Almeida, L. 2013. A Proof-Based Account of Legal Exceptions. *Oxford Journal of Legal Studies* 33(1): 133–68. https://doi.org/10.1093/ojls/gqs024.

Dung, P. M. 1995. On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and *N*–Person Games. *Artificial Intelligence* 77(2): 321–57. https://doi.org/10.1016/0004-3702(94)00041-X.

Dworkin, R. 1986. *Law's Empire*. Cambridge, MA: Belknap Press of Harvard University Press.

Dworkin, R. 1996. Objectivity and Truth: You'd Better Believe It. *Philosophy & Public Affairs* 25(2): 87–139. https://doi.org/10.1111/j.1088-4963.1996.tb00036.x.

Endicott, T. A. O. 2012. Legal Interpretation. In *The Routledge Companion to Philosophy of Law*. Ed. A. Marmor, 109–22. New York: Routledge.

Ferrer Beltran, J., and G. B. Ratti, eds. 2012. *The Logic of Legal Requirements: Essays on Defeasibility*. Oxford: Oxford University Press.

Gordon, T. F., and D. N. Walton. 2009. Legal Reasoning with Argumentation Schemes. In *Twelfth International Conference on Artificial Intelligence and Law* (ICAIL 2009), Barcelona, June 8–12, 2009, pp. 137–46. New York: Association for Computing Machinery. https://doi.org/10.1145/1568234.1568250.

Grice, P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hart, H. L. A. 1994. Postscript. In *The Concept of Law*. 2nd ed., 238–76. Oxford: Oxford University Press.

Horty, J. F. 2011. Rules and Reasons in the Theory of Precedent. *Legal Theory* 17(1): 1–33. https://doi.org/10.1017/S1352325211000036.

Horty, J. F. 2012. *Reasons as Defaults*. Oxford: Oxford University Press.

Horty, J. F., and T. J. M. Bench-Capon. 2012. A Factor-Based Definition of Precedential Constraint. *Artificial Intelligence and Law* 20: 181–214. https://doi.org/10.1007/s10506-012-9125-8.

Loui, R. P., and J. Norman. 1995. Rationales and Argument Moves. *Artificial Intelligence and Law* 3: 159–89. https://doi.org/10.1007/BF00872529.

MacCormick, N. 2005. Arguing about Interpretation. Chap. 7 in *Rhetoric and the Rule of Law: A Theory of Legal Reasoning*, 121–42. Oxford: Oxford University Press.

MacCormick, N., and R. S. Summers. 1991. Interpretation and Justification. Chap. 13 in *Interpreting Statutes: A Comparative Study*. Ed. N. MacCormick and R. S. Summers, 511–44. Aldershot, UK: Dartmouth Publishing Company.

Maranhão, J., and G. Sartor. 2019. Value Assessment and Revision in Legal Interpretation. In *Seventeenth International Conference on Artificial Intelligence and Law* (ICAIL 2019), Montreal, QC, June 17–21, 2019, pp. 219–223. New York: Association for Computing Machinery. https://doi.org/10.1145/3322640.3326709.

Marmor, A., and S. Soames, eds. 2011. *Philosophical Foundations of Language in the Law*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199572380.001.0001.

Modgil, S., and H. Prakken. 2013. A General Account of Argumentation with Preferences. *Artificial Intelligence* 195: 361–97. https://doi.org/10.1016/j.artint.2012.10.008.

Nair, S., and J. Horty. 2018. The Logic of Reasons. Chap. 3 in *The Oxford Handbook of Reasons and Normativity*. Ed. D. Star, 67–84. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199657889.013.4.

Patterson, D. 1996. *Law and Truth*. Oxford: Oxford University Press.

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

Pollock, J. L. 2008. Defeasible Reasoning. Chap. 23 in *Reasoning: Studies of Human Inference and Its Foundations*. Ed. J. E. Adler and L. J. Rips, 451–70. Cambridge: Cambridge University Press.

Prakken, H. 2010. On the Nature of Inference Schemes. In *Dialectics, Dialogue and Argumentation: An Examination of Douglas Walton's Theories of Reasoning and Argument*. Ed. C. Reed and C. W. Tindale, 167–85. London: College Publications.

Prakken, H. 2019. Modelling Accrual of Arguments in ASPIC+. In *Seventeenth International Conference on Artificial Intelligence and Law* (ICAIL 2019), Montreal, QC, June 17–21, 2019, pp. 103–12. New York: Association for Computing Machinery. https://doi.org/10.1145/3322640.3326703.

Prakken, H., and G. Sartor. 1998. Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law* 6: 231–87. https://doi.org/10.1023/A:1008278309945.

Prakken, H., and G. Sartor. 2009. A Logical Analysis of Burdens of Proof. In *Legal Evidence and Proof: Statistics, Stories, Logic*. Ed. H. Kaptein, H. Prakken, and B. Verheij, 223–53. Farnham, UK: Ashgate.

Prakken, H., and G. Sartor. 2015. Law and Logic: A Review from an Argumentation Perspective. *Artificial Intelligence* 227: 214–45. https://doi.org/10.1016/j.artint.2015.06.005.

Rahwan, I., and G. R. Simari, eds. 2009. *Argumentation in Artificial Intelligence*. Dordrecht: Springer.

Raz, J. 1975. *Practical Reason and Norms*. London: Hutchinson.

Raz, J. 1979. *The Authority of Law: Essays on Law and Morality*. Oxford: Clarendon.

Riveret, R., N. Oren, and G. Sartor. 2020. A Probabilistic Deontic Argumentation Framework. *International Journal of Approximate Reasoning* 126: 249–271. https://doi.org/10.1016/j.ijar.2020.08.012.

Rotolo, A., G. Governatori, and G. Sartor. 2015. Deontic Defeasible Reasoning in Legal Interpretation: Two Options for Modelling Interpretive Arguments. In *Fifteenth International Conference on Artificial Intelligence and Law* (ICAIL 2015), San Diego, CA, June 8–12, 2015, pp. 99–108. New York: Association for Computing Machinery. https://doi.org/10.1145/2746090.2746100.

Sartor, G. 2018. Defeasibility in Law. In *Handbook of Legal Reasoning and Argumentation*. Ed. G. Bongiovanni, G. Postema, A. Rotolo, G. Sartor, C. Valentini, and D. Walton, 315–64. Dordrecht: Springer. https://doi.org/10.1007/978-90-481-9452-0_12.

Sartor, G. 2022. Defeasibility and Burdens of Proof. *Rechtsphilosophie* 8(2): 152–66. https://doi.org/10.5771/2364-1355-2022-2-152.

Scalia, A., and B. A. Garner. 2012. *Reading Law: The Interpretation of Legal Texts*. St. Paul, MN: Thomson/West.

Soames, S. 2013. Deferentialism: A Post-originalist Theory of Legal Interpretation. *Fordham Law Review* 82(2): 597–613. https://ir.lawnet.fordham.edu/flr/vol82/iss2/9.

Solum, L. B. 2010. The Interpretation-Construction Distinction. *Constitutional Commentary* 27: 95–118. https://scholarship.law.georgetown.edu/facpub/676.

Tarello, G. 1980. *L'interpretazione della legge*. Milan: Giuffrè.

Toulmin, S. 2003. *The Uses of Argument*. Cambridge: Cambridge University Press. (1st ed. 1958.)

Walton, D., F. Macagno, and G. Sartor. 2021. *Statutory Interpretation: Pragmatics and Argumentation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108554572.

Walton, D., G. Sartor, and F. Macagno. 2016. An Argumentation Framework for Contested Cases of Statutory Interpretation. *Artificial Intelligence and Law* 24: 51–91. https://doi.org/10.1007/s10506-016-9179-0.

Walton, D., G. Sartor, and F. Macagno. 2018. Statutory Interpretation as Argumentation. In *Handbook of Legal Reasoning and Argumentation*. Ed. G. Bongiovanni, G. Postema, A. Rotolo, G. Sartor, C. Valentini, and D. Walton, 519–60. Dordrecht: Springer. https://doi.org/10.1007/978-90-481-9452-0_18.

