A logical analysis of argumentation in statutory interpretation

Giovanni Sartor
University of Bologna, Cirsfid
European University Institute, Law Department
Giovanni.sartor@unibo.it

Abstract

A logical model for arguments dealing with statutory interpretation will be here provided. The basic assumption is that interpretive arguments can be viewed as defeasible inferences: they support their conclusion, but this support is merely presumptive, since it may be challenged by counterarguments.

First a general pattern is introduced for representing arguments dealing with statutory interpretation. It is shown how interpretive arguments may be defeated by counterarguments, and how arguments and counterarguments may participate in larger argumentative interactions, where defeated arguments are reinstated when their defeaters are in turn defeated.

The idea is then developed of an interpretive argumentation basis, i.e., of a given set of interpretive canons and premises that can be used to build, or attack, interpretive arguments. The corresponding interpretive argumentation framework is considered, which includes the set of all interpretive arguments that can be constructed using the resources in the interpretive basis. Finally, a determination is made as to what claims are possibly (defensibly) or necessarily (justifiably) supported by a given interpretive framework.

1 Introduction

It scarcely bears mentioning that argumentation plays a key role in statutory interpretation: whenever a doubt is raised concerning the legal significance of a legal text, arguments are put forward to defend a certain interpretation or to challenge the alternatives to it. My focus is on explicit arguments meant to support or attack a determination of the meaning associated with a text, or at any rate its legal content. In order to capture all such arguments, I shall use the terms "interpretation" and "interpretive argument" in a very broad way. Thus, I shall not address the much-debated distinctions between understanding and interpretation (Dascal and Wroblewski 1988, Marmor 2005, Patterson 2005), or between interpretation and construction (Solum 2010) or the exercise of judicial discretion (Endicott 2012), or between identification and rectification of legal content (Soames 2013). On the one hand, direct/unreflected understanding falls outside the focus of explicit/deliberative/controversial interpretive processes; on the other hand, both interpretation in a strict sense and construction/rectification/completion of legislation are supported by arguments pertaining to the legal significance of textual documents (though

not all such arguments are based on reasons concerning the semantics or pragmatics of legislative language).

Legal theorists have proposed different classifications of interpretive arguments. For instance, the Italian legal theorist Tarello (1980) distinguishes the following 14 kinds of interpretive argument: a contrario, analogical, a fortiori, from completeness, from coherence, psychological, historical, apagogical, teleological, non-redundancy, authoritative, naturalistic, from equity, from general principles. MacCormick and Summers (1991) list 11 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 11 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, from honoured authority. Here I shall not discuss those lists (each of which has the merit of eliciting key patterns of legal reasoning), nor will I attempt an exhaustive classification of arguments, since my focus is instead on providing logical structures that can be shared by different kinds of interpretive arguments, namely, by interpretive arguments that appeal to different kinds of reasons (for a discussion see, Walton, Macagno and Sartor 2020, Ch. 1).

As a running example, I shall use the case of *Dunnachie v Kingston-upon-Hull City Council*, discussed by MacCormick (2005). This case concerns a claim of compensation for moral damages by an employee who had been unfairly dismissed, and as a result claimed to have suffered humiliation, injury to feelings, and distress. The key issues to be addressed in this case pertained to the interpretation of Section 123(1) of the UK Employment Rights Act 1996, which reads that "[T]he 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." In order to determine whether the claimant should be compensated not only for his financial losses, but also for his moral harm, the scope of the term "loss" had to be determined. The employee argued that an interpretation of this provision in the context of, and in coherence with, all the relevant sections of the statute, would grant him the recovery of *losses* other than financial losses narrowly construed. The employer argued that the relevant section of the current UK legislation 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 we are not presented with prevailing information against the argument. 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.

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 and does not cover moral harm. This shows that interpretive arguments can be defeated: it may be the case that the premises of an interpretive argument are sound, but that its conclusion nevertheless cannot be endorsed (since stronger, or at least not weaker, interpretive arguments support incompatible conclusions).

I shall refer in the following to some argument schemes —patterns for the construction of interpretive arguments— based on different interpretive canons. However, I shall first provide a general and most abstract pattern under which we can subsume all interpretive schemes, which we shall consistently use in the following to model interpretive debates.

The pattern is called *defeasible modus ponens*, and enables us to represent defeasible arguments in a form that mimics the *modus ponens* argument of deductive reasoning, namely, the inference pattern: if *P* then *Q*, *P*, therefore (deductively) *Q*. Indeed, a defeasible modus ponens inference has a similar form: if *P* then presumably *Q*, *P*, therefore (defeasibly) *Q*. The difference between deductive and defeasible modus ponens pertains to their conditional premise (their major premise). This premise is a truth-functional (or strict) conditional for deductive modus ponens: whenever the antecedent condition *P* is true, then necessarily the conclusion *Q* is also true. It is, on the contrary, a *presumptive* conditional that we have in the case of defeasible modus ponens: *P* only presumably or tentatively warrants *Q*, i.e., only as long as there are no prevailing reasons to the contrary (on defeasible reasoning in the law, see Sartor 2019).

In the following, I shall use a generalised version of defeasible modus ponens, where the antecedent is a general default which, using the terminology of Toulmin ([1958] 2003), we may call a *warrant*. The inference consists in matching the general 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. In Figure 1 you can see two of the best-known examples of defeasible inference, namely, the citizenship example by Toulmin ([1958] 2003, 94) and the ornithological example most widely used in the introduction to nonmonotonic logics. Each such inference has exceptions, and will indeed be defeated if any such exception obtains. For instance, the idea that Harry is a British citizen should be abandoned if we come to know that his parents are citizens of another country, as should the idea that Tweety flies if we come know that she is a penguin or an ostrich.

In interpretive reasoning, the warrant (major premise) is an interpretive canon, the minor premises are specific propositions matching the antecedent conditions of the canon, and the conclusion is a specification of the canon's consequent, as shown in Figure 1 (the oval contains "P" for "presumably"). By modelling different kinds of interpretive arguments as defeasible modus ponens inferences, based on different defeasible warrants, we represent all of them according to the same abstract logical pattern (see Prakken 2010b).

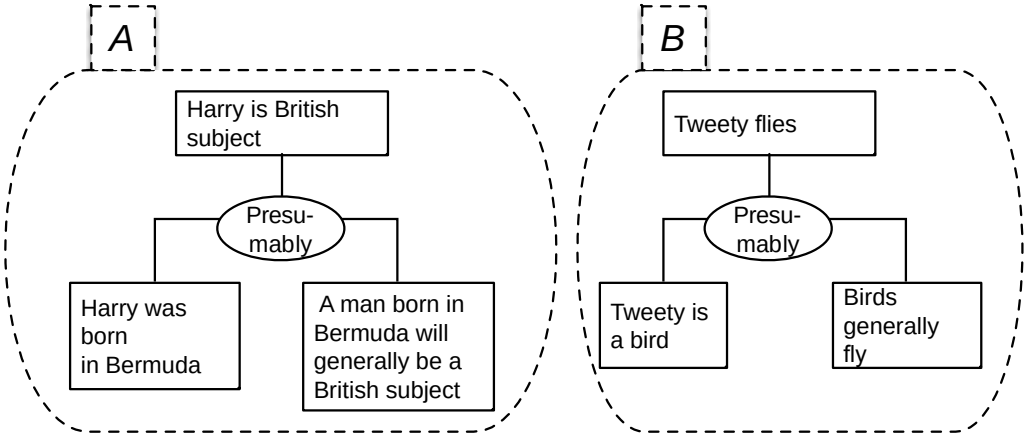

Figure 1. Defeasible modus ponens

3 The structure of interpretive arguments

In this section, I shall present a general structure for interpretive arguments. I shall first propose a canonical form for interpretive claims, and then a canonical way to support interpretive claims through the application of interpretive canons. I shall examine how interpretive considerations can be embedded in multi-step arguments so as to construct substantive arguments or derive legal interpretations or constructions based on the intention of the legislator, a contrario arguments, and analogical arguments. Finally, I shall address the justification of interpretive canons.

3.1 A canonical form for interpretive claims

The basic form for an interpretive claim is the assertion that an expression in a legal document should be interpreted as having a certain meaning.

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

  Here is an instance of this pattern:

- The expression "loss" in document "Employment Rights Act" should be interpreted as "pecuniary loss."

Interpretive conclusions can also be negative, i.e., they may claim that an expression should not be interpreted in a certain way.

- The expression "E" in document "D" should not be interpreted as "M."

Here is an instance of this pattern:

- The expression "loss" in document "Employment Rights Act" should not be interpreted as "pecuniary loss or injury to feelings."

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*."

Here are examples of both patterns:

- The expression "loss" in document "Employment Rights Act" should be interpreted as including "moral harm."
- The expression "loss" in document "Employment Rights Act" should be interpreted as excluding "moral harm."

These variations can be given a general scheme in the form

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

where I have included within square brackets the possible variants, separated by "|". By an expression *E* including a meaning *M*, I mean that the extension of *M* is a subset of the extension of *E* (all *M's* are *E's*). By an expression *E* excluding a meaning *M*, I mean that the extension of *M* is included in the complement of *M* (no *M's* are *E's*). Obvioulsy, the negation of an inclusion statement does not logically entail the corresponding exclusion statement: it may be the case that some *M*'s are not *E'*s and at the same time some are. However in natural language the proposition "the *M's* are not included in the *E's* "' would usually be understood as expressing (or at least implicating) the exclusionary claim that all *M*'s are not *E*'s, rather than the weaker claim that there exist at least one *M* which is not an *E*. For instance, the claim "merely moral harms are not included in compensable losses" would be understood as asserting that no merely moral harm is a compensable loss, rather than as the weaker claim that there exist one instance of merely moral harm which is not compensable, while others can be.

3.2 A canonical form for interpretive arguments

Let us now consider arguments linking reasons (minor premises) and warrants (major premise) to interpretive conclusions. We start with a "recommending" example of the argument from Ordinary Language, namely, with the recommendation to adopt a certain interpretation since it fits that canon.

1. Interpreting the expression "loss" in document "ERA" as having the meaning "pecuniary deprivation" fits the canon of Ordinary Language (minor premise).

2. If interpreting an expression “E” in a document “D” as having a certain 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.”

Using the diagrammatic form introduced above we obtain the following diagram:

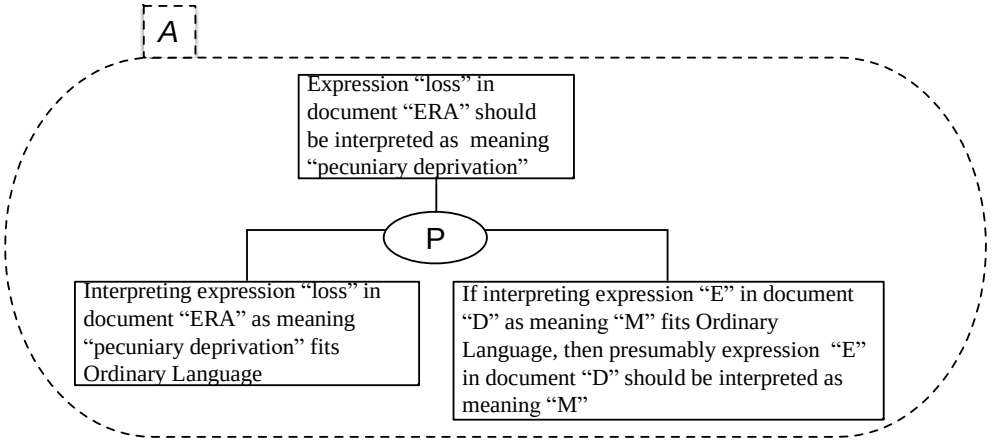

Figure 2. Interpretation according to Ordinary Language

This pattern can be used to support the inclusion of a certain class in the meaning of the interpreted text according to the canon of Purposiveness, as in the example in Figure 3.

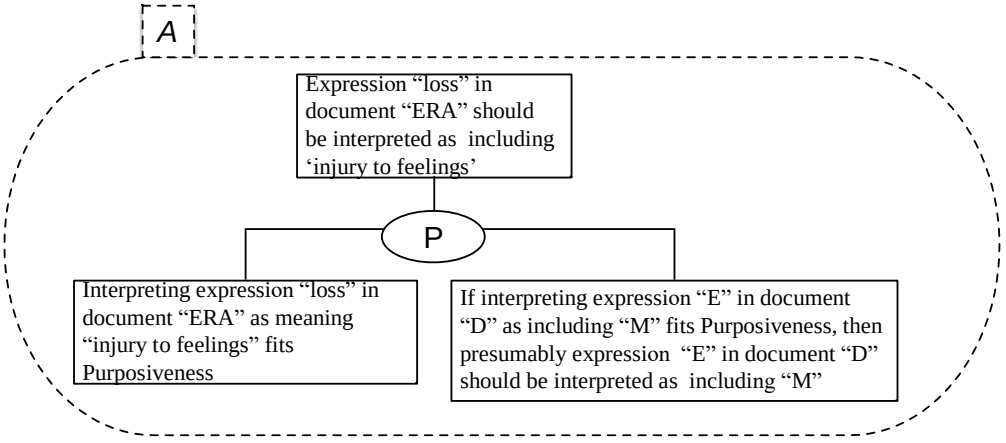

Figure 3. Inclusionary interpretation

3.3 Multi-step interpretive arguments

In the examples above one-step interpretive arguments have been presented, claiming that an interpretation should be adopted since that the ascription of a certain meaning fits a certain requirement (corresponding to Ordinary Language, Purposiveness, etc.). In order to expand interpretive arguments, we have to combine the argument claiming that an

interpretation should be adopted since it fits the requirement of an argument scheme, with further arguments arguing that the proposed interpretation indeed satisfies that requirement.

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

1. The expression "loss" is usually understood by English speakers as meaning "pecuniary loss" (minor premise).

2. If the expression "E" is usually understood by English speakers as meaning "M," then interpreting expression "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 multi-step argument shown in Figure 4.

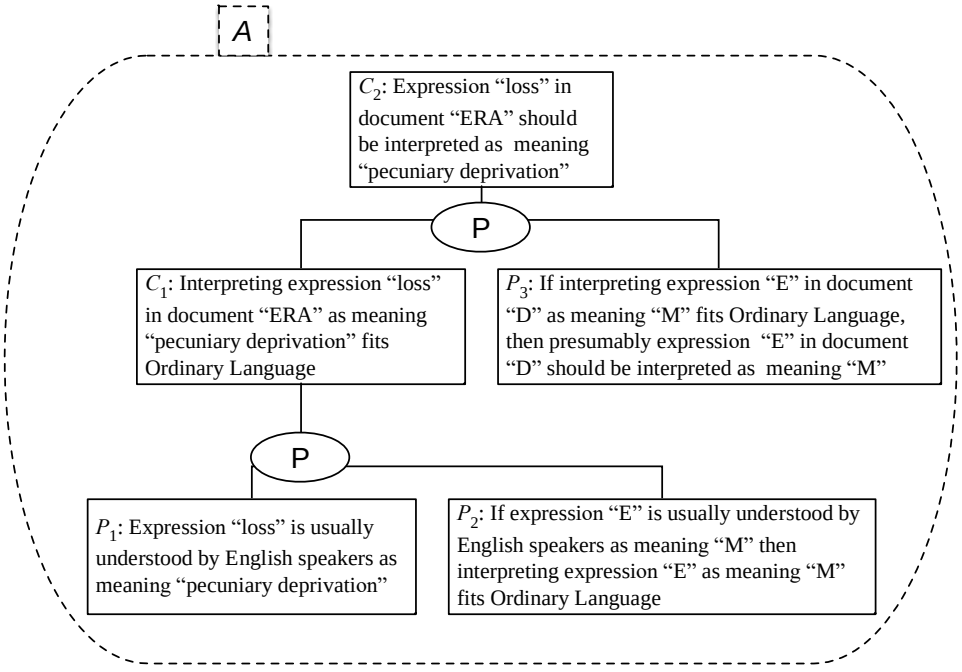

Figure 4. Multi-step interpretive argument

Note that the largest argument *A*, culminating in conclusion C_2, includes the subargument with conclusion C_1 (which works as a premise for the last inference step in the argument. Each premise P_1, P_2, P_3 can also be viewed as an argument, namely, as a most basic kind of argument, which only consist in asserting a claim without any supporting reason. It could perhaps be said that there is an implicit reason supporting such claims, namely, the very fact that they were uttered, under the assumption that people usually assert what they believe to be true (and that for the most part they are correct).

3.4 Compressing interpretive arguments

The above presented structure of interpretive arguments can be compressed, using a label to indicate the argument being used, and leaving implicit the standardised top part of the argument, namely, the inference step that leads from the fact that an argument fits a scheme to the interpretive claim that, based on that argument, a certain interpretation should or should not be adopted (on the compression of arguments, see Loui and Norman 1995).
In Figure 5 you can see, for instance, an affirmative interpretive scheme concerning Ordinary Language. Though a compressed form may better suit a synthetic and readable presentation of arguments, here I shall use the uncompressed form, which is better suited to my analytical goals.

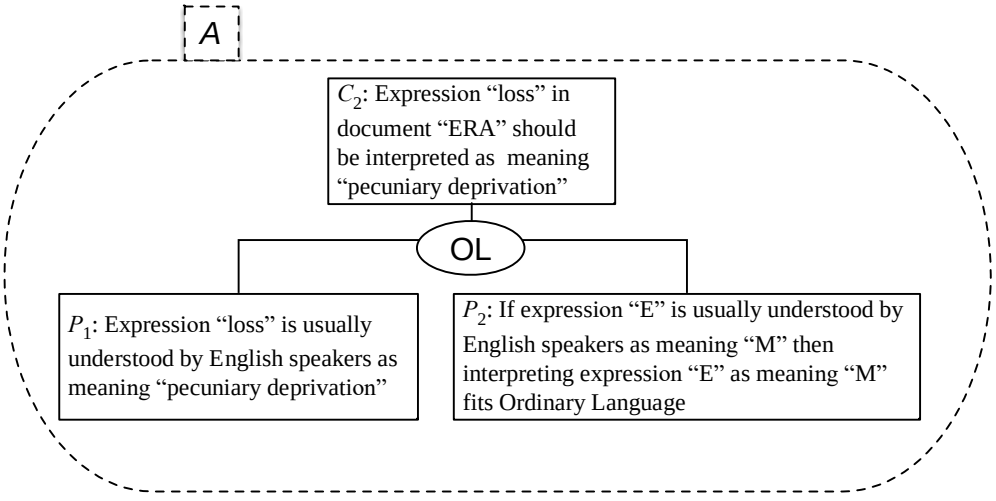

Figure 5. Compressed interpretive argument

3.5 The intention of the legislator: A transcategorial argument

As we observed in the Introduction, arguments concerning the legislator's intention may be viewed as transcategorial, i.e., they may be based on further interpretive arguments. In the framework here proposed, this can modelled by combining two arguments: (a) the argument that an interpretation should be adopted since it corresponds to the legislator's intention, and (b) the argument that this interpretation corresponds to that intention, since the same intention can be reconstructed by applying a certain interpretive scheme to the legislator's expression.

Here (Figure 6) is an example of a transcategorial intention-based argument based on Ordinary Language. Note that the box with thicker border contains the claim that the proposed interpretation (interpreting "loss" as pecuniary deprivation) fits Ordinary Language, but this claim is not used to conclude directly for the adoption of this interpretation. It is instead used to support the claim that this interpretation fits legislative intent.

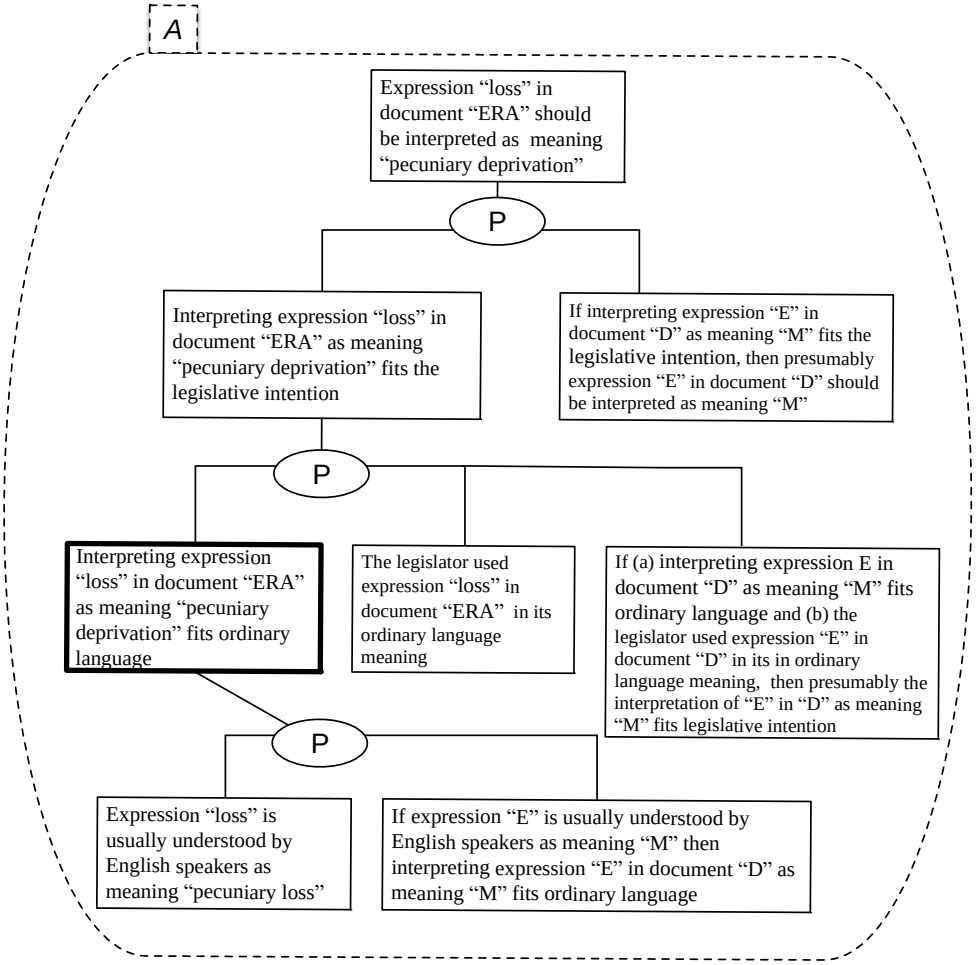

Figure 6. Transcategorial intention-based argument according to Ordinary Language

A similar argument, pointing to the opposite conclusion, namely, that the "loss" should be interpreted as also including injury to feelings, can be constructed by assuming that the legislator's intention is captured through a teleological interpretation, as shown in Figure 7. Note that the intention-based arguments may also not be transcategorial, which is the case when they are based on direct factual evidence of the legislator's intention (e.g., on political declarations or legislative materials).

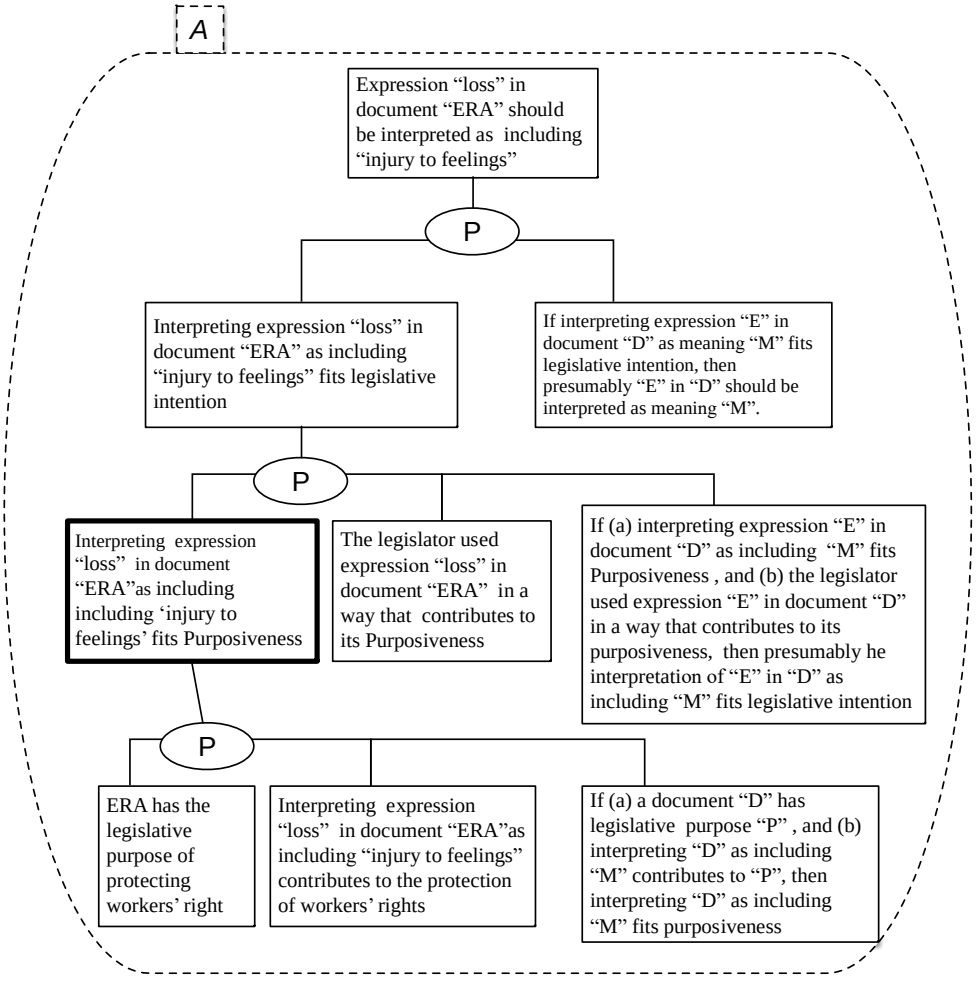

Figure 7. Transcategorial intention-based argument according to Purposiveness

3.6 Interpretive arguments and substantive legal conclusions

As observed in the Introduction, interpretive arguments may be embedded in legal arguments that address substantive legal issues. The example in Figure 8 shows an argument culminating in the conclusion of a substantive liability rule, in which the rule is supported by an interpretive argument, while the rule’s factual antecedent may result from arguments based on the available evidence.

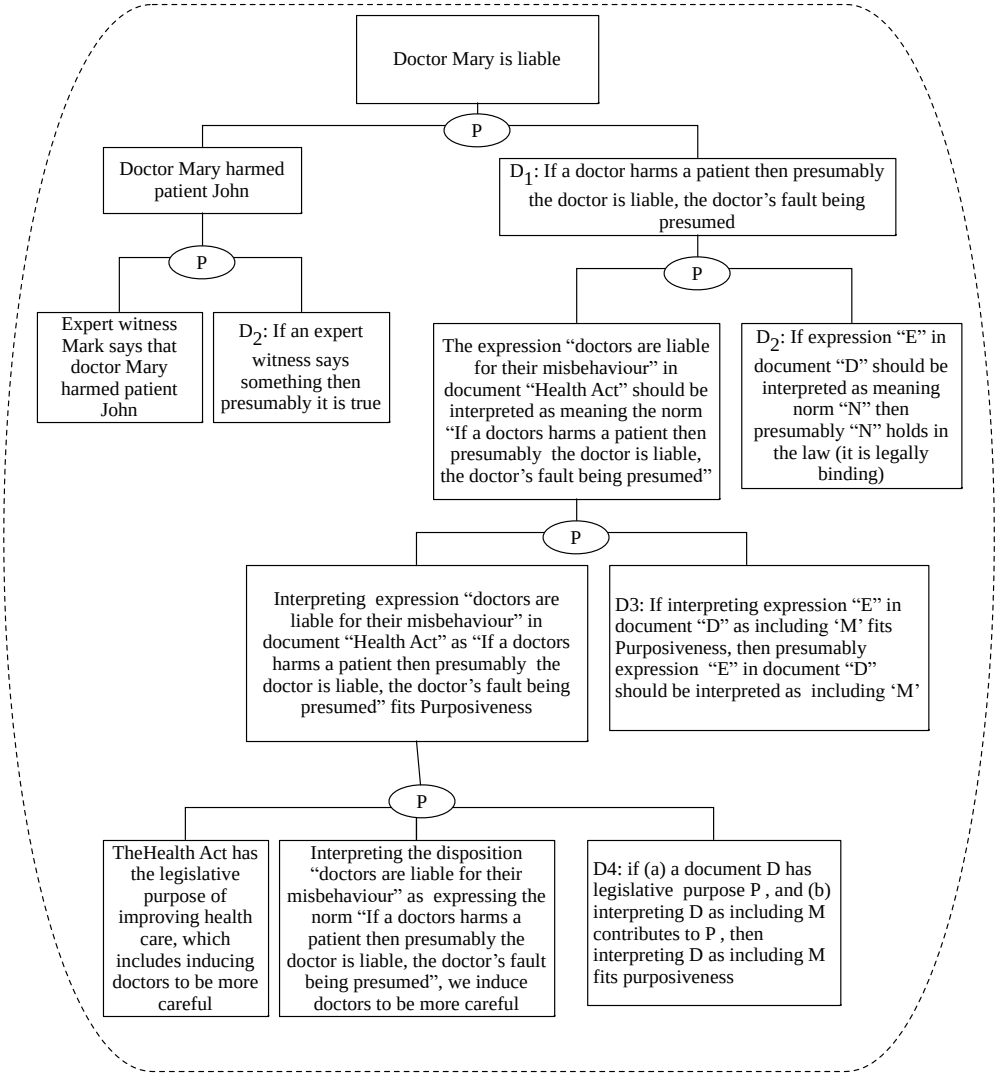

Figure 8. Multi-step legal argument

More precisely, this argument includes all of the following:

- A norm-based argument that Mary is liable, since she harmed her patient and doctors are liable for harm done to their patients, unless they are shown not to be at fault (the normative premise).
- A teleological interpretive argument: the provision in the Health Act on doctors' liability must be interpreted in this way, since this interpretation contributes to increasing diligence in the medical profession, thus contributing to Purposiveness.
- An empirical argument based on an expert's testimony supporting the conclusion that there was a causal link between Mary's behaviour and the patient's harm.

This argument is subject to a series of possible attacks against each of its elements: the top inference may be challenged by defeating the presumption of fault, but arguing that Mary was not at fault (she used the available medical knowledge correctly); the interpretive subargument can be attacked by contesting the very idea that the proposed interpretation,

by establishing a presumption of medical fault, promotes careful behaviour among doctors (on the contrary, this presumption may undermine patient care, since doctors may become too risk-averse, knowing that they may face the difficult task of proving a negative fact, namely, that they did *not* act negligently); the empirical subargument can be rebutted by providing a contrary expert opinion, or it can be undercut by challenging the expert's reliability, among other options (on the transition from interpretive arguments to substantive conclusions, see .

3.7 A contrario, analogical, and a fortiori arguments

An *a contrario argument* is based on, but is not limited to, the exclusionary use of an interpretive canon (see Canale and Tuzet 2008). As we shall see in the following, an *a contrario* argument —as the opposite argument based on legislative analogy— includes three steps. The first step usually consists in applying canons based on linguistic meaning (e.g., on Ordinary Language or on Technical Language) so as to conclude that the expression at issue should be interpreted as excluding a certain class. The second step (usually implicit) consists in arguing that the class, being excluded from the expression being interpreted, fails to match the operative facts specified in the provision in which that expression is contained. The third and decisive step consist in arguing that the negation of the provision's conclusion applies to members of the excluded class, since they do not match the provision's operative facts.

Let us provide a simple example to clarify this point. Consider, in front of a restaurant, a sign that reads "Dogs are allowed," and consider what on that basis can be argue with regard to cats. An *a contrario* argument to the effect that cats are not allowed may start with determining that the term "dogs" should be interpreted as excluding cats. This conclusion may be provided by considering that since in Ordinary Language the term "dog" does not cover cats, we should interpret it as excluding cats from its scope (an Ordinary Language argument).

This inference, however, still does not deliver the conclusion that cats are not allowed (that they are prohibited). The second step consists in arguing that as a consequence of excluding cats from the meaning of "dogs," cats fail to meet the provision's antecedent condition (its operative facts) allowing dogs in, a provision whose normative content may be expressed, in a conditional form, as follows: "If you are a dog, then you are allowed in." (Note that we are assuming that the strong or prescriptive negation of a permission is a prohibition, as distinguished from the descriptive normative proposition saying that the given norms entail no permission; see Alchourron 1969.) The third step consist in claiming that, since cats do not meet this provision's antecedent condition, the complementary normative position (prohibition rather than permission) applies to them (see Figure 9).

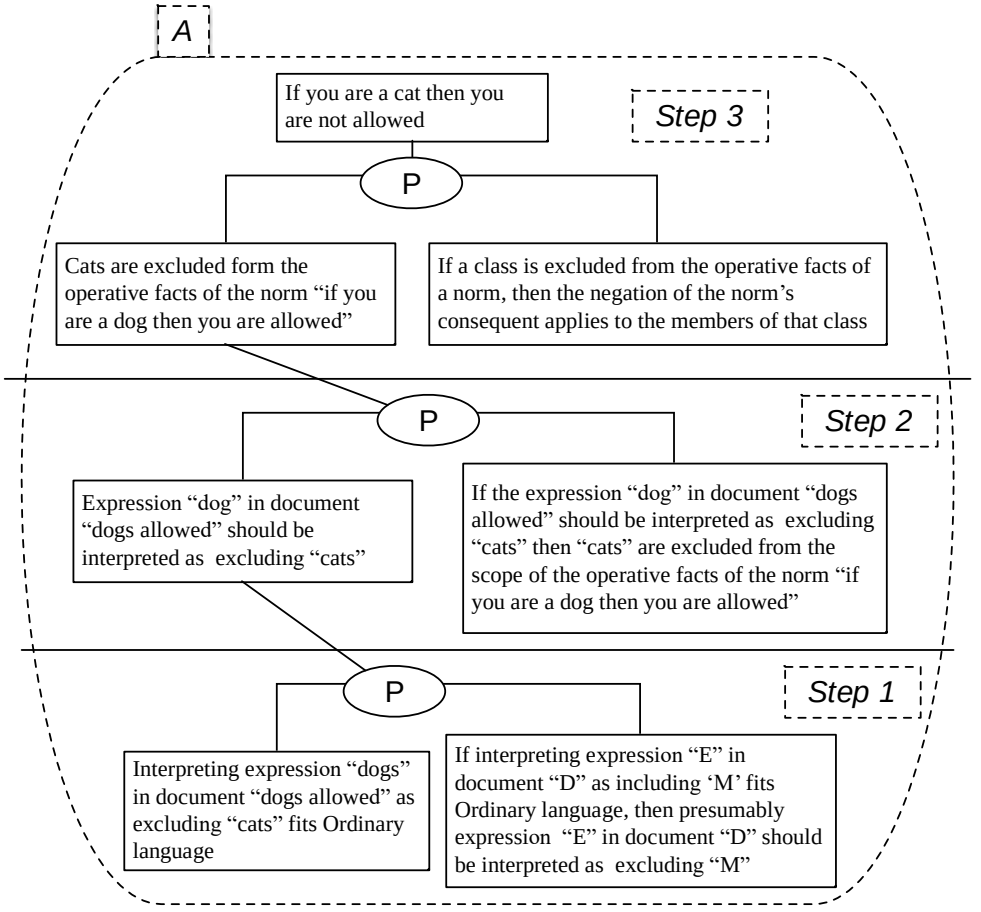

Figure 9. A contrario implicature

The *a contrario* argument scheme may be linked to the idea of an general exclusionary norm, originally advanced by Zitelmann (1903), under which each legal system contains a general principle according to no criminal or civil sanction should be imposed on an action (the action is therefore permissible) unless the sanction is set forth in a legal norm. As framed here, however, the principle has a broader scope, applying to each single legal rule, and it is defeasible. According to it, the mere fact that a certain class is not included (it is excluded) from the operative facts of a particular legal norm (or, more generally, from a set of norms sharing the same conclusion) may ground the presumption that whatever legal effect may be established by that norm (by such norms) does not apply to the excluded class.

A contrario arguments that deny a legal property based on excluding a class from an expression's linguistic meaning are also viewed as application of the canon *expressio unius* (for: *expressio unius est exclusio alterius*, i.e., "the expression of one thing is the exclusion of the other"). As an example of the application of this canon, Scalia and Gardner (2012) consider a New Hampshire statute that shielded municipalities from "damages arising from insufficiencies or hazards on public highways, bridges, or sidewalks . . . when such hazards are caused solely by snow, ice, or other inclement weather." By applying the *expressio unius* canon, the competent court refused to interpret this rule as also covering public parking lots. Consequently, it held a municipality responsible for the harm suffered by person who fell on ice on a public parking lot.

3.8 A contrario arguments and statutory analogies

The *a contrario* argument —denying that a legal property holds for a certain class since that class is excluded from a norm's operative facts— is defeated if additional norms can be found that explicitly ascribe the property to that class. Consider again the *a contrario* argument presented above, according to which cats not allowed since the relevant provision (the sign in front of the restaurant) only addresses dogs. Suppose that the restaurant owner in the example, responding to a series of complaints by cat lovers, puts up an additional sign on the door reading "Cats allowed." In the context of this further sign, the *a contrario* implicature that cats are not allowed no longer holds. On the other hand, based on the combination of the "Dogs allowed" and "Cats allowed" signs, *a contrario* arguments can be mounted for animals different from both cats and dogs.

The results of *a contrario* arguments can be contradicted by the use of analogies. For instance, based on the permission for dogs, it could be argued that cats are likewise allowed (rather than being *a contrario* forbidden), given that the two animal species are relevantly similar (Figure 10).

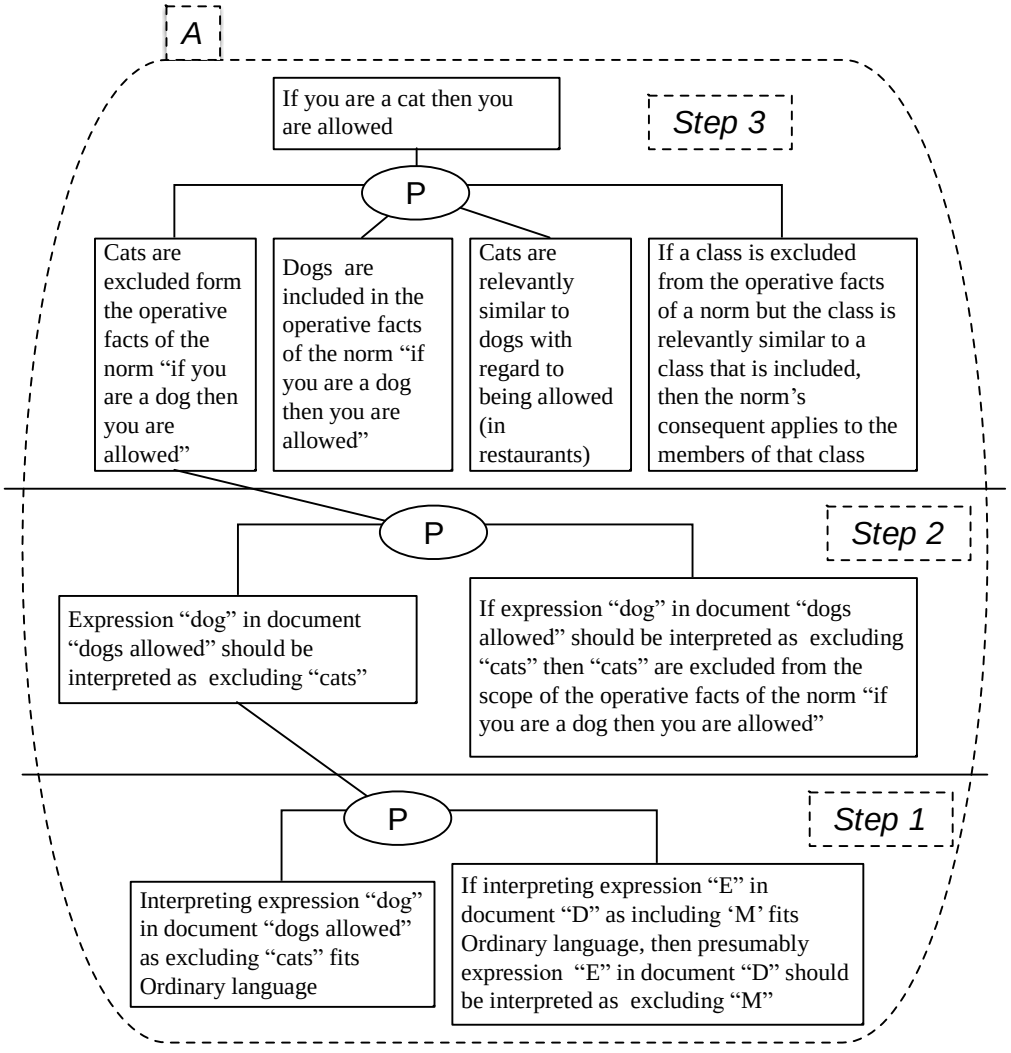

Figure 10. Analogical reasoning

In an analogy, the legal discipline established by a norm for a certain class *A* is applied to another class *B* based on the fact that *B* is relevantly similar to *A*. The relevant similarity may be based on the fact that the two classes are in included in more general class *C*, and that *A* and *B* share features that, relatively to entities of type *C* favour the legal discipline at stake. For instance, to support the relevant similarity of dog and cats, with regard to access to the restaurant, it may be claimed that both are animals such they provide company to their owners and can usually be kept under control in such a way that nuisance to other patrons is avoided. According to the Italian legal theorist Norberto Bobbio ([1958–1960] 1993), the analogy scheme —according to which the legal outcome provided for certain cases can be extended to similar cases— can be construed under the label of a "general inclusionary norm," which conflicts with the "general exclusionary norm" provided by the *a contrario* scheme. For Bobbio, the clash between these two general norms —a clash for which no abstract solution is available— gives rise to the indeterminacy of legal systems.

3.9 Analogies, extensive interpretations, and a fortiori reasoning

Statutory analogies and extensive interpretations are contiguous: both argue that a legal provision applies to facts that do not fall under that provision's Ordinary Language interpretation, and both can be used to attack *a contrario* arguments leading to the opposite conclusion. However, there is the need to draw a distinction between the two arguments schemes, especially to the extent that analogies in statutory interpretation are prohibited in certain domains of the law —such as criminal law, where the principle of legality requires the law to pre-define crimes and sanctions. The difference between the two schemes consists in the fact that a statutory analogy accepts that the expression being interpreted *excludes* the class that should be regulated in the same way, whereas extensive interpretation argues that this expression *includes* that class. For instance, an analogy against the previous *a contrario* argument denying immunity to New Hampshire municipalities would accept that parking lots are not included in the meaning (the extension) of the expression "public highways, bridges, or sidewalks." It would, however, claim that parking lots are so similar to the classes of objects covered by that expression (e.g., since parking lots are also designed for vehicles and support public mobility) that we should assume that, relative to parking lots, municipalities enjoy the same immunities they enjoy relative to public highways, bridges, or sidewalks. An extensive interpretation of "public highways, bridges, or sidewalks" would on the contrary reject the exclusion of parking lots from that expression. It would claim that parking lots are indeed included in the meaning (extension) of that expression (e.g., since a public parking lot can be viewed as part of the highways to which they are connected), which directly leads to the application of the immunity to parking lots.

On this analysis of *a contrario* and analogy arguments in legal interpretation, we may wonder whether these arguments fit into the concept of statutory interpretation, strictly understood as the determination of the meaning of a legislative expression, since both arguments lead to normative conclusions concerning facts that are assumed to be excluded from the meaning of the interpreted provision. However, *a contrario* and analogical

arguments may still be viewed as interpretive implicatures of the provisions concerned, which are pragmatically linked to the interpretation of such provisions.

Similar considerations also apply to *a fortiori* arguments. Such arguments presuppose that a norm exists that regulates a certain class *A* of entities or situations, e.g., prohibiting or permitting certain actions to be performed by (or with regard to) such entities in such situations. Typically, an argument a fortiori claims that another class *B* of entities or situations should be regulated in the same way, on the basis of the fact that such a class requires even more (*a fortiori ratione*, literally, from a stronger reason) the same regulation. The stronger reason may consist in the fact that both classes *A* and *B* are included in a superclass *C*, and *B* possesses additional factors favouring the same regulation relatively to entities of type *C* or lacks factors disfavouring it relatively to such entities. It may also consist in the fact that *B* possesses at a larger degree some scalable dimensions proportionally favouring that regulation or possesses at smaller degree some scalable dimensions that would disfavour it (on factors and dimensions, see Bench-Capon and Rissland 2002). In the example above, it could be argued that since dogs are explicitly allowed in a restaurant, cats should also be allowed in *a fortiori* since they are animals that fit such a liberal entry regime even more than dogs. The greater suitability of cats may be established by arguing that

- animals ought to be allowed in restaurants (*i*) the more their company is important to their owners (a scalable favouring dimension), and (*ii*) the less they may be a nuisance to other patrons (a scalable disfavouring dimension); and
- cats are (*i*) equally important to their owners as dogs are, and (*ii*) less of a nuisance to other patrons.

If we accept these (highly debatable) premises we may indeed conclude that cats ought to be allowed entry on even stronger grounds than dogs (cats are indeed Pareto-superior to dogs when it comes to the benefit of allowing them in restaurants, since they are equally good under one standard, and better under the other). An *a fortiori* argument may be directed at the legislator (or the restaurant owner) to support the request that a new provision be enacted (granting access to cats as well in the future). However, the argument may also consist in a legal construction based on the existing legal provisions (the sign "Dogs are allowed") from which a directly applicable conclusion (Cats are also allowed) is obtained.

3.10 The rationale behind interpretive canons

In all the examples so far presented, argumentation has concerned the application of an interpretive canon, but we have not considered whether an interpretive canon in itself may be supported by arguments, possibly to respond to challenges raised against it. Can we develop arguments that support the use of an interpretive canon, namely, rationales for the use of the canon, or do interpretive canons escape rational debate, not needing any justification?

The issue of the rationale of reasoning schemes has been addressed in general terms in Walton and Sartor (2013), where it is argued that reasoning schemes may be subject to pragmatic justifications, pointing to how well they serve the needs of the practical or

epistemic activity in which they are used. This idea also applies to the interpretive canons used in legal reasoning, and in fact it would not be difficult to find rationales justifying each of them. For instance, it may be argued that the canon of Ordinary Language is supported by the need to provide legal certainty, protect the justified expectations of citizens, and limit the arbitrary power of legal decision-makers. Similarly, it may be argued that the canon of Purposefulness is supported by the need to achieve the (socially useful) goals pursed by a (democratic) legislator, and that the canon of Coherence is needed to achieve useful outcomes when implementing different laws and to provide citizens with consistent. These rationale-based arguments could be seen instances of the so-called argument on good consequences, as shown in

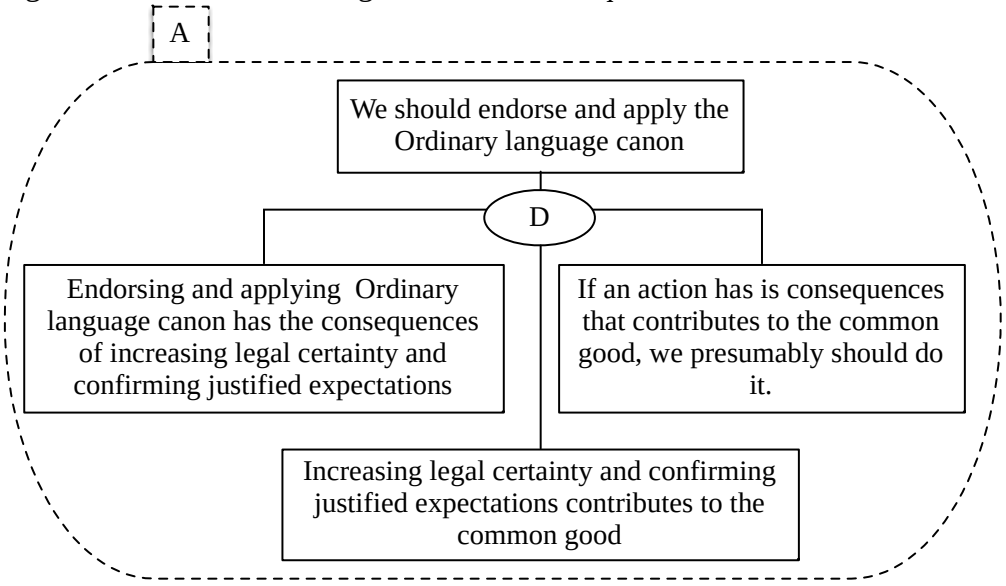

Figure 11.

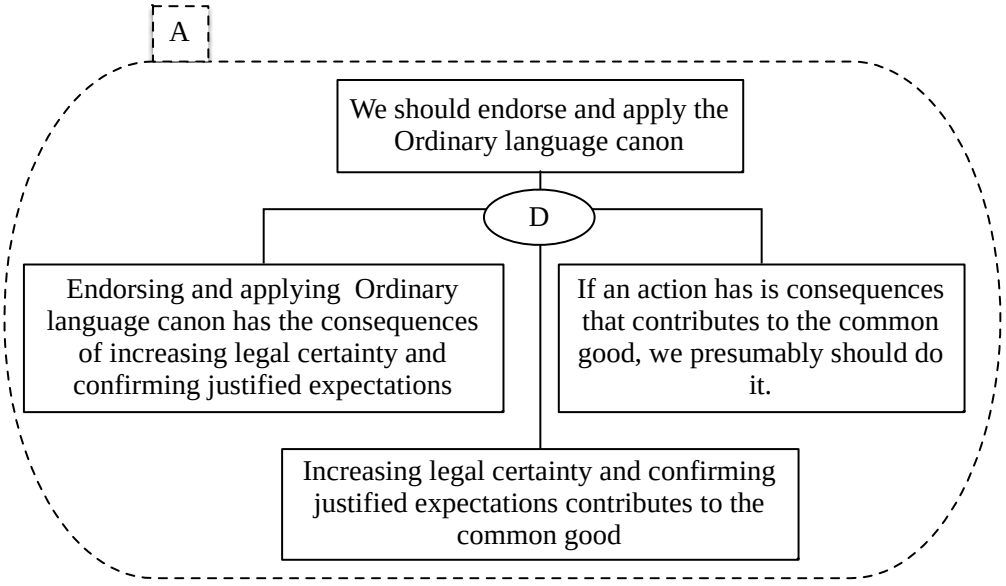

Figure 11. A justification for an interpretive canon

A justification that applies to all interpretive canons, regardless of their specific content, consists in the very fact of their existence as shared reasoning tools, to be presented and

accepted as reasons for interpretation. For one thing, the fact that interpretive canons are shared in the legal community motivates individual reasoners to use them as useful tools by which to support their claims. And, for another, the fact that interpretive canons are shared makes them socially useful, since their shared use facilitates legal interactions and ensures some degree of convergence.

This does not make the substantive rationales for each canon irrelevant to legal argument. In particular, these rationales become relevant when different canons lead to opposite outcomes, since the preference for one canon over another may be adjudicated according to the importance of the goals (values) supporting the application of each canon and the extent to which such goals may be affected by applying or not applying that canon in the case being considered (e.g., how much is legal certainty important in comparison to substantive goals, and to what extent may legal certainty be promoted by interpretations conforming to Ordinary Language, or negatively affected by interpretations departing from it?).

4 Conflicts between interpretive arguments

In this section, conflicts between interpretive arguments shall 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 arguments when at least one of these arguments 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 *B*'s conclusion (or the conclusion of a subargument of *B*), i.e., if *A*'s conclusion is incompatible with a conclusion established by *B* (or by a subargument of *B*, as a preliminary step to its final conclusion).
- 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.

Figure 12 and Figure 13 exemplify the two types of attack between competing arguments in the context of legal interpretation:

- Arguments *A* and *B* in in Figure 12 rebut each other, since the first concludes that "loss" in "ERA" should be interpreted as "pecuniary detriment" and the second concludes that it should be interpreted to also include "injury to feelings" (two incompatible conclusions).
- Argument *B* in Figure 13 undercuts argument *A*, since *A* reaches a conclusion using teleological reasoning, and *B* argues that a teleological inference does 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, an assumption that in its own turn can obviously be attacked, on nontextualist grounds).

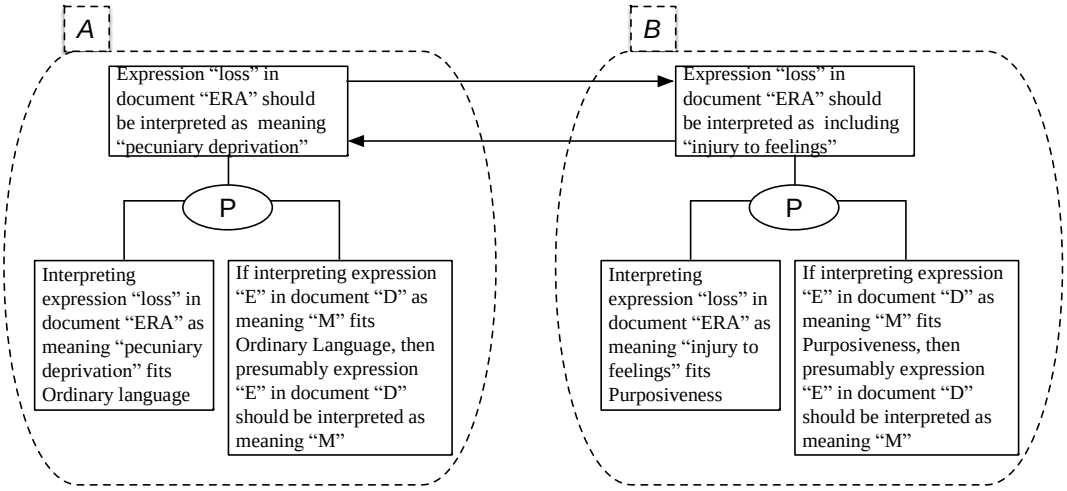

Figure 12. Rebuttal in interpretation

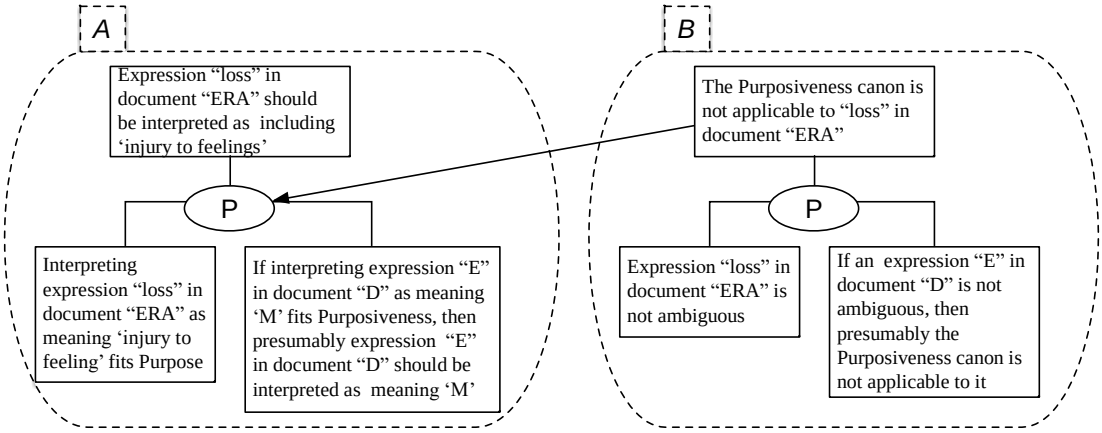

Figure 13. Undercutting in interpretation

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 in by most English speakers. This argument could be undermined by producing some evidence (e.g., a dictionary) according to which "loss" is usually understood, by most English speakers, as limited to certain nonmonetary detriments (e.g., permanent bodily harm that does not affect earning capacity and does not require additional therapy).

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 turn, is attacked by another argument *C*.

In clarifying this point, it is useful to specify a formal semantics for arguments, namely, to spell out the precise conditions which an argument in an argument set should meet in order to be acceptable relative to that set. The most influential formal semantics for arguments has been proposed by Dung (1985) 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 different approach, one that for our purposes is equivalent to Dung's semantics (see Baroni, Caminada, and Giacomin 2011) but is instead based on labelling all arguments in the given argument set with one of two labels: IN (acceptable) or OUT (inacceptable). The basic idea is that an argument is IN if it that has no attackers that is IN: only an attacker which is IN can revert to OUT the argument it attacks; 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 14, which combines the arguments in Figure 10 and Figure 12. Argument *A* is IN even if attacked by *B*, since *B* is OUT, being attacked by *C*, which is IN, having no attacker. Therefore, on the basis of all the arguments in Figure 14 we should conclude that "loss" should indeed be interpreted as a pecuniary deprivation, since this follows from the IN-argument *A*.

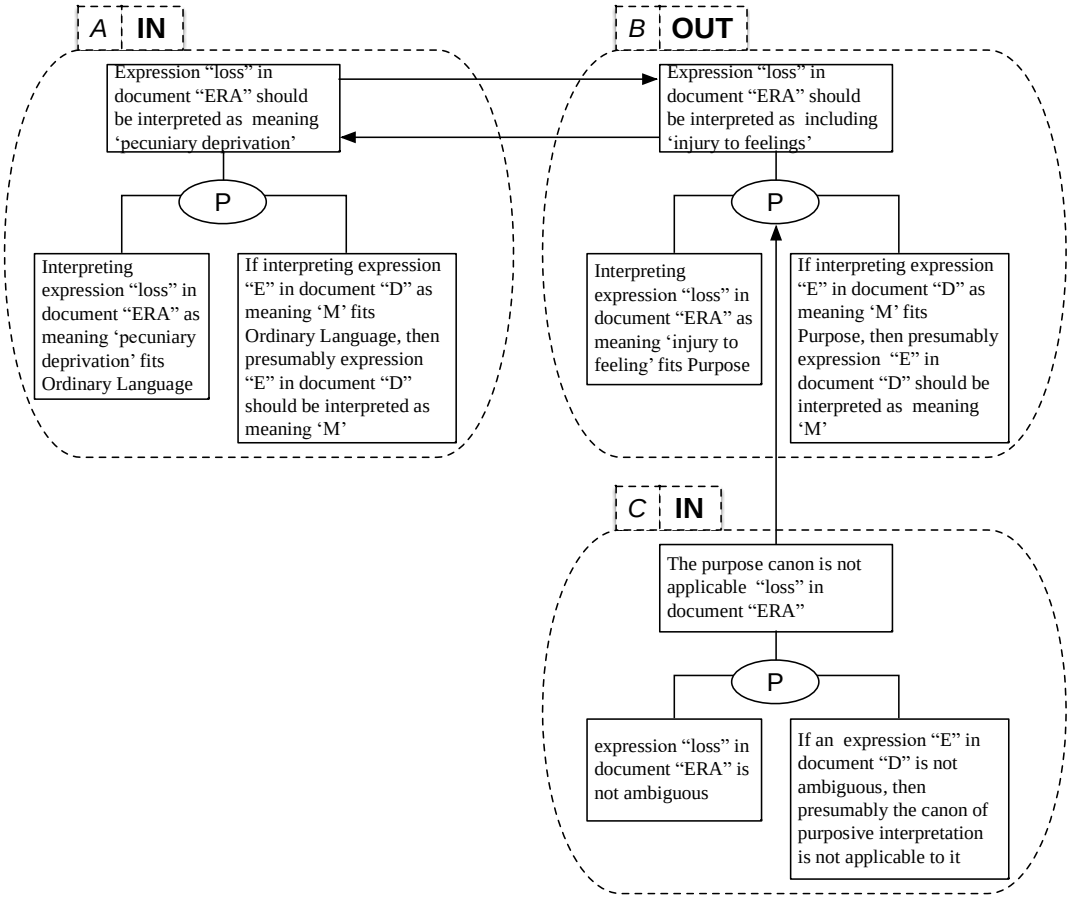

Figure 14. Labelling of arguments, reinstatement

4.3 Undecided conflicts

Our analysis of the interaction of arguments must be completed with a discussion of the case in which two arguments rebut each other and it is not apparent which of two should prevail. When empirical information is key, this issue can be addressed by relying on the burden of proof: if there is a burden of proof regarding an operative fact, and there are equally strong arguments for and against the existence of that fact, then the legal effect conditional on that fact cannot be triggered, not even defensibly. Assume, for instance, that the plaintiff has the burden of proving negligence to establish liability in a tort case: the plaintiff then will lose (the judge will not find the plaintiff 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 providing 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 (on the burden of proof, see Prakken and Sartor 2006).

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 is indeterminate, unless the prevalence of such argument can be

established or assumed. For instance, going back to the case of 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 (Non-Redundancy);
- 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 (Intention of the legislator).

In the context of the theory of argumentation here developed multiple pairs of 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), then in each pair of incompatible interpretive arguments, each argument successfully attacks (defeats) the other. Consequently, neither is sufficiently supported.

Let us consider a conflict between the Ordinary Language interpretation and a purposeful interpretation, pointing to opposite conclusions, as in Figure 12, and assume that the no undercutter is available. We will then have obtained the situation represented in Figure 15.

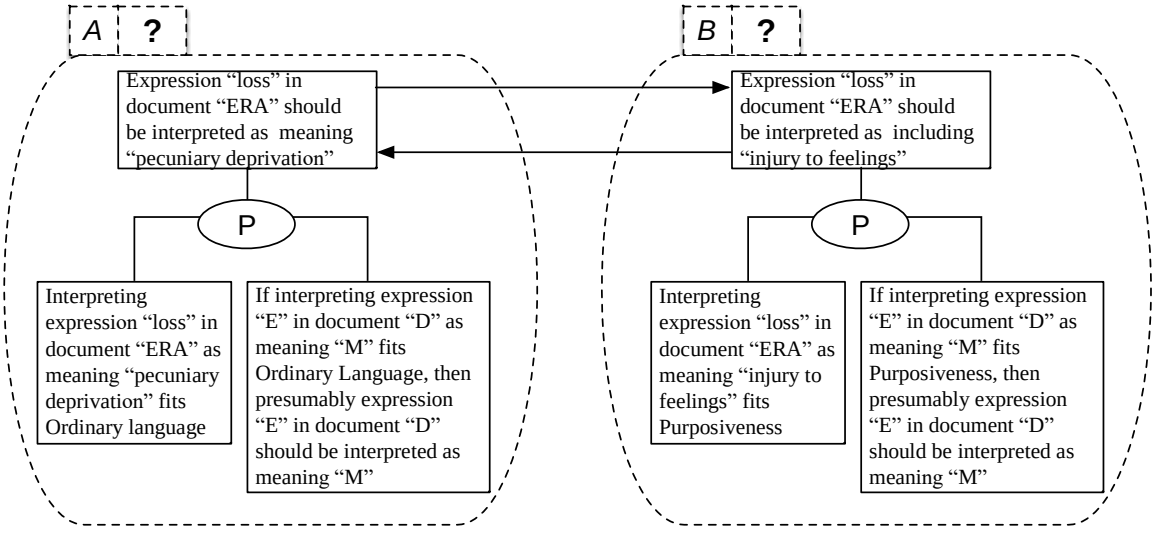

Figure 15. Undecided conflict

Given the information represented in the argument graph in Figure 15, the outcome of the conflict of 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. Thus, our argument set appears to be indeterminate, i.e., unable to provide a single reliable outcome. Incompatible alternative choices seem possible (we can assume that either of the two argument prevails), so that it may be argued that either outcome is compatible with the law (or that neither of them is required by it).

4.4 Priority arguments

In the last section we noted that incompatible interpretive argument may be constructed by using accepted interpretive canons. We also observed that the choice of one such argument is in a way self-validating: if we choose to say that one argument is IN, then the other argument is OUT, and vice versa. This perspective seems consonant with some realist views on legal interpretation, such as that advanced by Bobbitt (1991), according to which the choice of one interpretation over the other is a political or moral choice, which is not regulated by law, and is left to the decision-maker's conscience (or in any case, choice).

However, this need not always be the case, since it is often the case that the conflict between two arguments that rebut one another (as in Figure 15) may also be addressed though argumentation: it may be possible to provide further arguments that adjudicate the conflict, namely, preference arguments that support the prevalence of one argument over the other (Prakken and Sartor 1997, Modgil and Prakken 2013). To the extent that the criteria for such a choice are provided by law, we can still say that the law calls for single solution even when alternative interpretive arguments are available.

Preference arguments may 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's limit case), or in a vested argument, where the preference claim is supported by reasons. Consider, for instance, the argument set presented in Figure 16. This argument set includes, besides the mutually rebutting arguments *A* and *B*, an additional argument *C* that adjudicates the conflict between arguments *A* and *B*. In particular, *C* argues that the Purposiveness argument *B* prevails over the Ordinary Language argument *A*, since in this case the goal pursued pertains to constitutional values (and the pursuit of constitutional values prevails over Ordinary Language considerations).

According to *C*, *B* has the upper hand over *A*. This means that —according to C— *B*, being stronger, successfully attacks (i.e., defeats) *A*, while *A*'s attack against *C* is not successful. Thus, *C*'s impact on the conflict between as *A* and *B* is that, based on *C*, the prevailed-upon argument *A* fails to successfully attack *B*. 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).

To take into account the possibility that attack links are also attacked, we need to rewrite as follows the previous rules on the assignment of IN/OUT states to arguments:

- 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 16 shows how a priority argument can affect the IN/OUT states of the arguments whose conflicts it adjudicates. The priority argument *C* states that Purposiveness prevails over Ordinary language, so that argument *B* prevails over argument *A*. Doing that, *C* successfully attacks the attack-link from the weaker argument (*A*) to the stronger argument (*B*), which is consequently turned to 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.

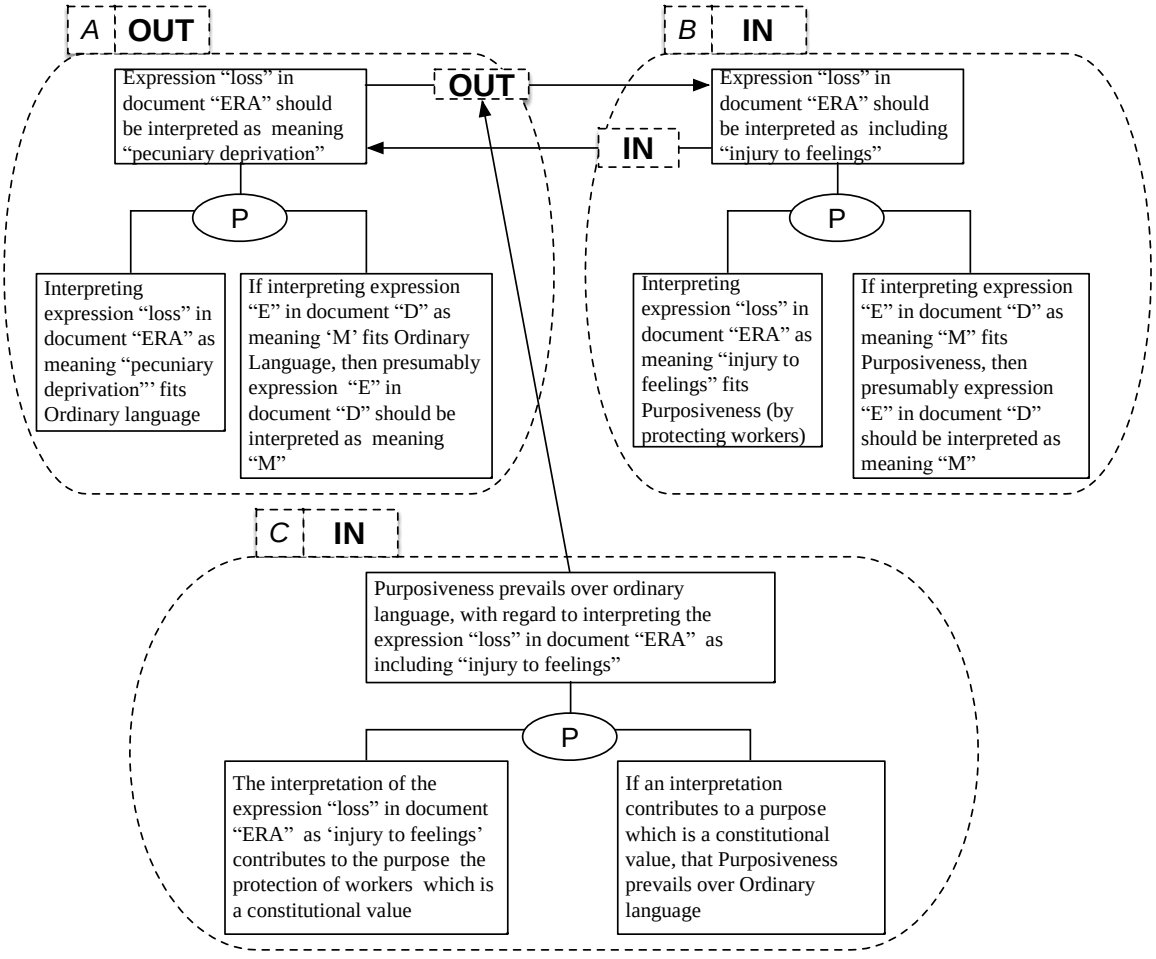

Figure 16. Solution of an interpretive conflict through a priority argument

4.5 Statuses of arguments and conclusions

Considering all the possible assignments of IN/OUT labels to the arguments and attack-links at stake —consistently with rules (1) and (2) above (and also taking into account the possibility that attack relations over rebuttals are also attacked)— we can classify the arguments in an argument set into three separate jointly exhaustive sets:

- Arguments that are *justified*. These are arguments that are IN according to every correct assignment of IN/OUT labels, i.e., every assignment that complies with the previously stated rules.
- Arguments that are *overruled*. These are arguments that that are OUT according to every correct assignment.
- Arguments that are *merely defensible* (defensible but not justified). These are arguments which are neither justified nor overruled: they are IN according to some correct assignments and OUT according to other correct assignments.

The IN/OUT status of a merely *defensible argument* depends, directly or indirectly, on an unresolved rebutting conflict, i.e., a conflict in which incompatible arguments rebut each other within an argument set that does not determine which of the mutually rebutting arguments prevails.

On the basis of the status of an argument we can determine the status of the conclusions it supports. First of all, if a justified argument supports a conclusion, then that conclusion is itself justified. with regard to the conclusions of merely defensible arguments, however, we may wonder: Certainly, the conclusions of such arguments are themselves defensible, but under certain conditions we may rather consider such conclusion as being justified. This happens when multiple alternative defensible arguments are available and they all point to the same conclusion. For instance, assume that it is undetermined whether "loss" is to be interpreted as "pecuniary detriment" or as "pecuniary detriment and harm to feelings," but that an unfairly dismissed worker only asks to be compensated for the monetary loss he has suffered. In this situation we can certainly claim that this conclusion is legally supported, since all defensible interpretations point to it, without the need to choose among them.

More precisely, on this approach an interpretive conclusion can be said to be *justified* if there is at least one IN argument supporting that conclusion in every correct IN/OUT assignment (although, depending on how the assignment is made, different arguments supporting that conclusion may be considered IN). Similarly, an interpretive conclusion can be said to be *overruled* if all the arguments supporting that conclusion are OUT in every correct IN/OUT assignment. Finally, an interpretive conclusion is merely defensible if there is at least one IN argument supporting that conclusion on some, but not in all, correct IN/OUT assignments.

Note that our definition for overruled conclusions covers both conclusions that are supported by arguments, all of these arguments being overruled, and conclusions that are not supported by any argument. This, too, is a distinction that can be introduced if it proves useful to the analysis.

4.6 The accrual of convergent arguments

An argumentative canon can subsume other canons, and multiple canons can point to the same conclusion: in so doing, these multiple canons may "accrue," i.e., reinforce their conclusions (on accrual, see Prakken 2019).

A multi-step argument for an interpretive conclusion may include convergent arguments in favour of applying a single interpretive canon. For instance, a supporter of Scalia's interpretation —the view that the right to bear arms also involves the use of

weapons for personal defence— could argue that this interpretation should be adopted on the grounds 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.

An argument for a certain interpretation could also be supported by the convergent application of multiple canons. For instance, the view that "loss" should also include injuries to feeling 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 counterargument, argument having the same conclusion would converge and indeed accrue: the combination of the separate arguments supporting the same conclusion would have a greater force than each of them can carry separately.

Besides convergences and conflicts between alternative canons, there may also be conflict 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 17, 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).

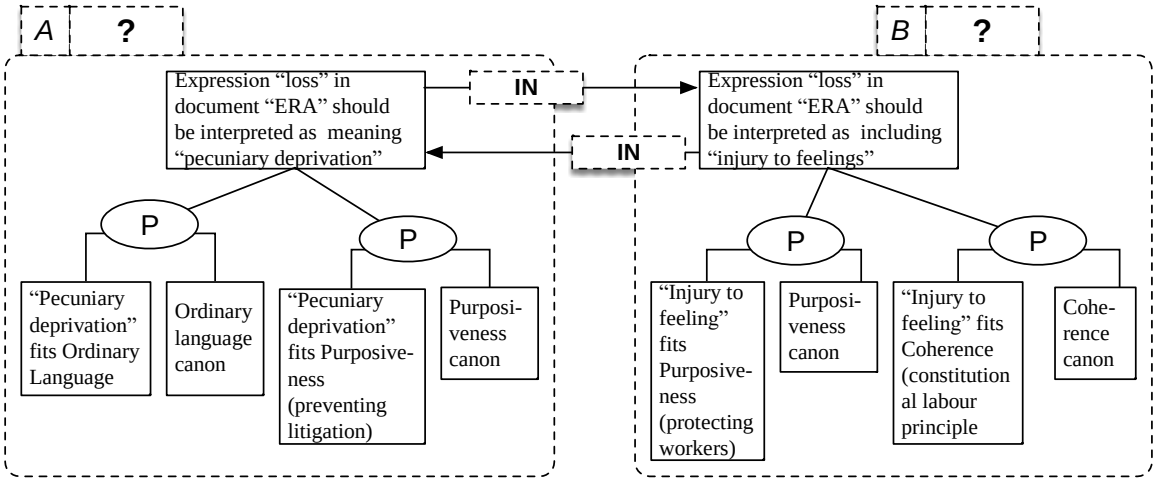

Figure 17. Accrual of arguments

4.7 Interpretive canons (and relevant facts) as argumentation bases

We have so far considered arguments and their interactions, i.e., argument conflicts giving rise to attacks. Let us now look at the premise sets that provide the ingredients for constructing a set of interacting arguments, i.e., in our case, the set of the available interpretive canons and the facts matching such canons.

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 counterarguments. It is an *argumentation basis*, in the sense of a knowledge base (a set of premises) that can be used for constructing an *argumentation framework* (a set of interacting arguments).

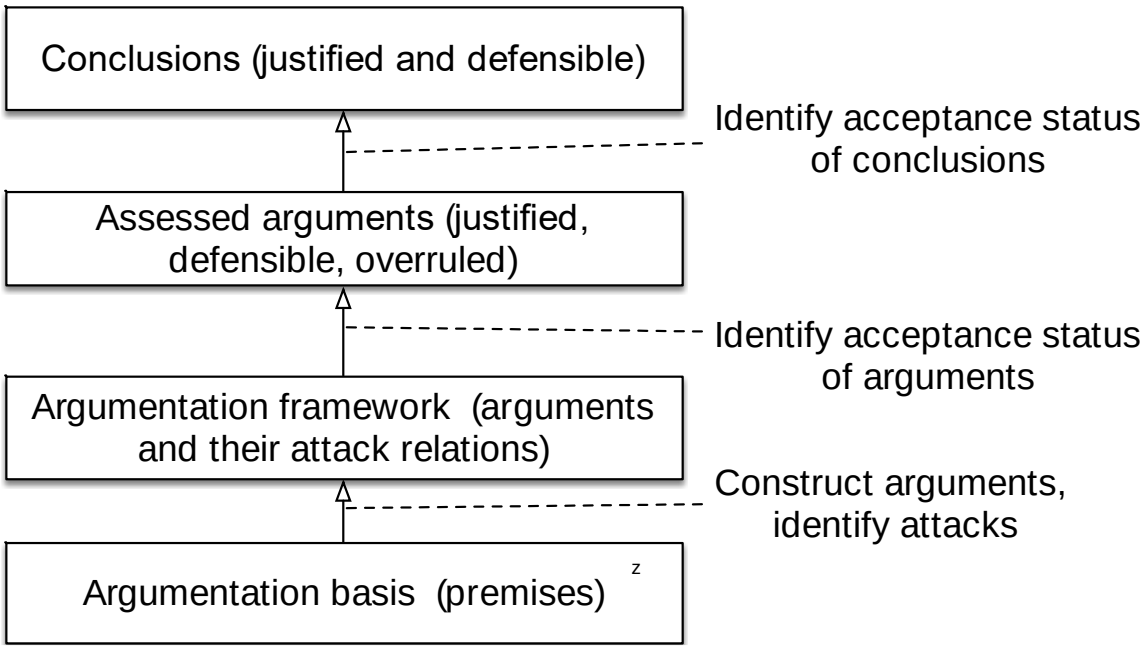

Figure 18. Argumentation basis

Figure 18 (adapted from Baroni, Caminada, and Giacomin 2011) illustrates the process for determining the inferential semantics of an argumentation basis —a set of premises to be used in argumentation— namely, the set of all conclusions that are justified or at least defensible relative to that basis. First, we construct the argumentation framework resulting from the argumentation basis, i.e., we construct all arguments that can be obtained by using the premises in the basis, and we identify the attack relations between such arguments. Then we determine what arguments and attack links 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 overruled relative to the given argumentation basis. Finally, we identify the status of the *conclusions* of these arguments. Figure 19 applies to legal interpretation the general idea represented in Figure 19.

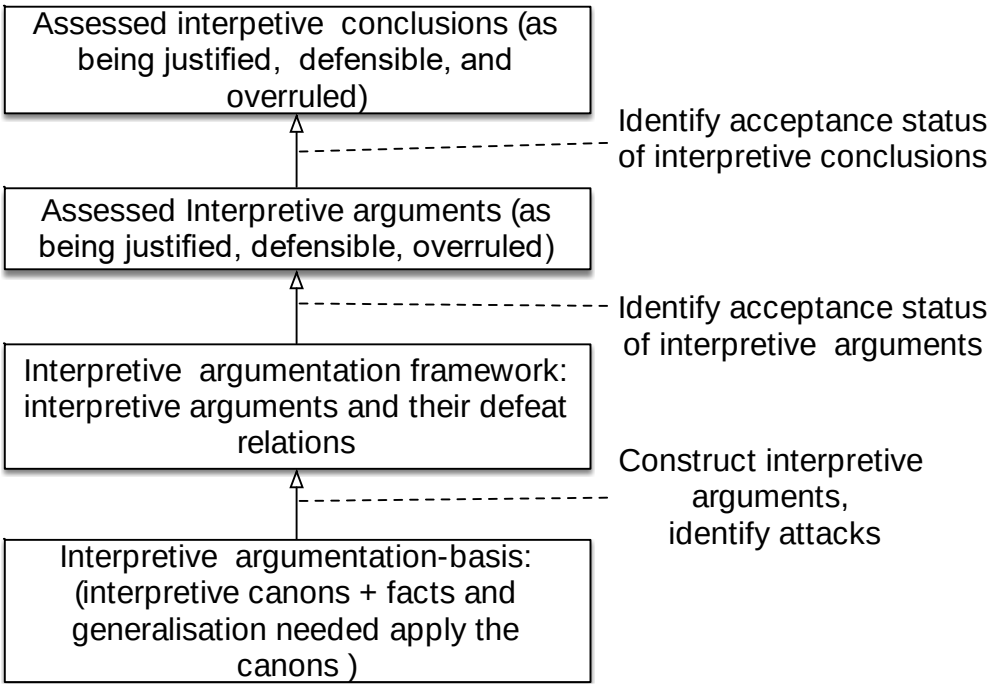

Figure 19. Interpretive argumentation basis

4.8 From the interpretive argumentation basis to the interpretive argumentation framework

An interpretive argumentation basis —the set of interpretive canons, coupled with the textual content of the documents to which the canons are applied, along with any further premises that may be relevant to the interpretation of such texts— determines an argumentation framework, which may include multiple interpretive arguments, some of which may be incompatible. 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 preventing litigation. In view of the goal of protecting workers, however, Purposiveness favours, by contrast, the inclusion of "injury to feeling," 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 the legal system makes alternative preferences available concerning which of these canons prevails (a range of preferences that may depend on the different ideological positions of judges and legal scholars). We then have the situation is depicted in Figure 20. The IN/OUT status of all arguments 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.

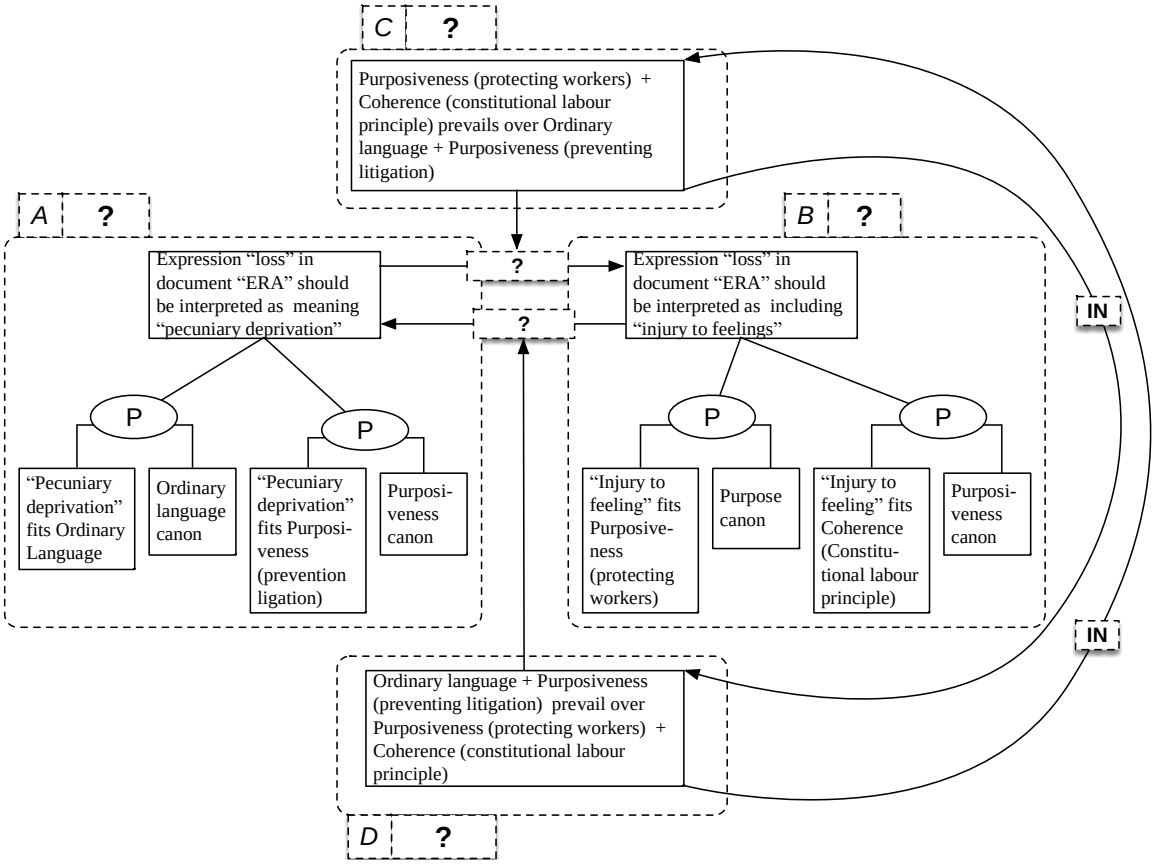

Figure 20. An interpretive framework allowing for alternative defensible solutions

Assume, in view of the problem of addressing this matter, that a legal system only offers the premises for building these interpretive arguments, and assume in particular that this system does not provide indications 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 , and so are their incompatible conclusions. We can also describe this situation by saying that both arguments are "legally possible" and neither is "legally mandated." In fact, the available information does not enable us to decide whether or not to follow the interpretation that (through the application of the canons of Purposiveness and Coherence) favours the substantive constitutional values at stake over the interpretations that (through the application of the canons of Ordinary Language and Purposiveness) favour the "formal" values of certainty and the prevention of litigation.

The situation changes if the argumentation basis is expanded with a meta-preference (a preference between preferences) for *C* over *D* as depicted in Figure 21. Such a meta-preference can be the conclusion of a vested argument, e.g., we could argue that interpretations favouring substantive constitutional values prevail since this has been established by a precedent of the constitutional court.

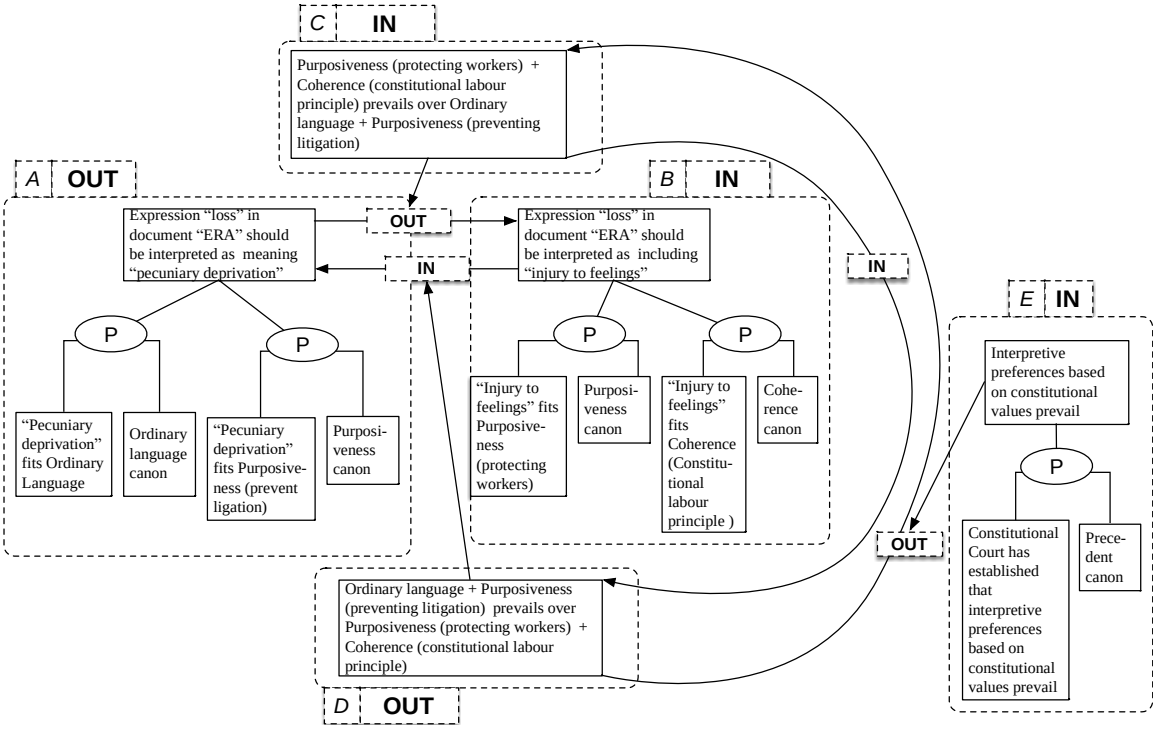

Figure 21. An interpretive framework with a single justified solution

The interpretive framework represented Figure 21 has a justified interpretive solution, namely, the conclusion of argument *B*, which is indeed justified in that interpretive framework (while *A* is now overruled). This follows from the fact that according to *E* (which is IN, having no attacker), the weaker preference argument *D* is unable to successfully attack *C*. Thus, *C* is definitely IN, and consequently *A* is considered to be weaker than *B*. Consequently, *A* is OUT, being successfully attacked (defeated) by *B*.

4.9 Is there always one right answer?

Our analysis enables us to provide a fresh approach to the classic issue of whether and under what conditions there is a single legal right answer to an interpretive issue (e.g., the issue on how the term "loss" in the Employment Rights Act.

This issue can have a precise solution when the question concerns a circumscribed argumentation basis and corresponding argumentation framework. Under such conditions, there will be a single right answer to an interpretive question if that answer is a justified conclusion of that framework. Conversely, there will be multiple equally right answers if the argumentation framework provides multiple merely defensible answers, namely, alternative defensible conclusion. In the first case, the single justified conclusion answering the issues can be said to be a necessary interpretation, relative to the given framework, necessity consisting in the fact that under all correct assignments of IN/OUT labels, there is an IN argument with that conclusion. In the second case each defensible conclusion answering the issue may be said to merely possible interpretation, relative to the given framework, mere possibility consisting in the fact that each answer is supported by IN arguments under *some* assignments, but not all.

Thus, for instance, relative to the argumentation framework in Figure 20, alternative possible (defensible) answers can be said to 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 21, there is a *single* necessary (justified) answer to that question.

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 the law is assumed to deliver a single justified solution or multiple defensible ones depends on two concurring factors: what existing legal sources say (what their content is), and what counts as a determinant of the law.

The first factor concerns the extent to which the socially available legal sources — broadly understood to also include shared interpretations schemes and legal principles— provide criteria for deciding conflicts among alternative canons. For instance, a legal system that requires a deferential approach to interpretation, giving a preference to Ordinary Language and other linguistic factors, except under exceptional conditions, would restrict defensible interpretations, and expand justified ones. On the contrary, the range of merely defensible interpretations would be expanded, and the range of justified interpretations correspondingly restricted, to the extent that priorities among different canons are unavailable or are in conflict.

The second criterion pertains to what, according a legal theory, is considered a determinant of the law. If we assume that political morality also contributes to shaping or determining the law, and that there is a single correct political morality (as argued, for instance, by Dworkin 1985 and Alexy 1989), then the range of defensible interpretations could be in principle be limited, since moral arguments ought to solve most, if not all, conflicts between interpretive arguments. Going back to our example concerning the interpretation of "loss," if the correct moral reading of the Constitution would indicate a prevailing preference for interpretations favourable to workers, on the ground of their dignity and equality, then the indeterminacy in Figure 20 could be solved on this basis. However, different people may disagree about the content that should be ascribed to correct morality —e.g., some may reject the understanding of dignity and equality leading to that conclusion— or they may disagree about whether substantive principles of political morality should prevail over ideas of legal certainty and formal equality. Their different moral views may lead them to give different contents to the interpretive argumentation basis they endorse, and then to different views on what the law entails: in the event of such a disagreement, different people could endorse different interpretive argumentation bases and could therefore 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 defensible interpretations, legal analysis should recognise the indeterminacy of the law. This limitation of legal reasoning, however, is compatible with the law allowing, or even mandating, the competent decision-maker to

select one of the available defensible solutions, drawing on nonlegal, moral or political, considerations (see Hart 1984).

Finally, note that the framework here proposed allows for a distinction to be made between interpretive claims, on the one hand, and assertions about interpretive bases, on the other —a distinction that parallels the well-known distinction between non-truth-functional norms and truth-functional normative propositions (see Alchourron 1969). Interpretive claims, here modelled in the form "Expression *E* in document *D* should be interpreted as *M*," may be viewed as prescriptive assertions, expressing recommendations for the interpretive community, rather than as truth-functional propositions. On the contrary the metalevel assertion that a certain interpretive claim is a justified, defensible, or overruled conclusion relative to a certain interpretive basis can, in our framework, be viewed as a truth-functional proposition, albeit one that only concerns the given interpretive basis. The truth conditions of such interpretive propositions are given by the previously discussed criteria for IN/OUT assignments.

However, the assertion that a certain claim is justified or defensible relative to an interpretive framework *N* may cease to be merely descriptive when the content of *N* is unspecified. In this case, the claim may presuppose (and implicitly argue for) the inclusion of interpretive recommendations or preferences in the concerned interpretive framework, in such a way as to make the claim true. For instance, assume that a it is claimed that in UK law "loss" means "pecuniary deprivation" (i.e., that this claim is justified relative to the interpretive basis provided by UK law, and to the relevant facts), without relativizing this claim to a particular set of canons and other premises (which may include canons and premises supporting counterarguments). Then this claim may implicitly advocate that the interpretive recommendations/preferences supporting that claim should be included in UK law. In fact, the assertion of an interpretive proposition —to the effect that an interpretive claim is justified or defensible relative to a certain legal system— involves two distinct assertions: the assertion that the interpretive claim is indeed justified or defensible relative to a certain interpretive basis (as identified by the interpreter) and the assertion that this interpretive basis faithfully mirrors the relevant content of the legal system being considered. Both assertions can be challenged: the first on the basis of logical analysis (following the model here presented), the second on the basis of empirical information, and possibly of moral claims (depending on whether it is assumed that the law also excludes or includes certain moral premises, either contingently or necessarily).

5 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 logical framework here presented draws inspiration from the ASPIC system for structured argumentation (Prakken 2010, Modgil and Prakken 2013), which develops ideas indroduced in Dung (1995) and Prakken and Sartor (1997). While a number of formal models exist for with arguing cases, making analogies and distinction (Ashley 1990, Prakken Sartor, 1997, Horty 2011, Horty, and Bench-Capon 2012), fewer contributions address statutory interpretation through formal argumentation (see Araszkiewicz 2013, Rotolo, Governatori and Sartor 2015, Da Costa et al. 2017).

It is important to note that the semiformal framework here provided covers the phenomenon of 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 mechanism that are involved in the application of interpretive argument schemes, which involve logical, linguistic, pragmatic, decision-theoretical, and other considerations (some of these aspects are addressed in Walton, Macagno, and Sartor 2018; Macagno, Walton, and Sartor 2019; and Maranhao and Sartor 2019; see also Brewer 2011 and Marmor and Soames 2011, Walton, Macagno and Sartor 2020).

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.

6 References

Araszkiewicz, M. (2013). Towards systematic research on statutory interpretation in ai and law. In *Proceeding of the 2013 conference on Legal Knowledge and Information Systems: JURIX 2013*, pp. 15 – 24. IOS.

Ashley, K. D. (1990). Modeling Legal Argument: Reasoning with Cases and Hypotheticals. MIT.

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

Bobbio, N. ([1958-1960] 1993). *Teoria generale del diritto*. Giappichelli.

Brewer, S. (2011). Logocratic method and the analysis of arguments in evidence. Law, Probability and Risk, 10:175–202.

Canale, D. and G. Tuzet (2008). On the contrary: Inferential analysis and ontological assumptions of the A contrario argument. *Informal Logic 18*, 31–43.

Dascal, M. and Wroblewski, J. (1988). Transparency and doubt: Understanding and interpretation in pragmatics and in law. Law and Philosophy, 7:203–24.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmono- tonic reasoning, logic programming, and n–person games. Artificial Intelligence, 77:321–57.

Gordon, T. and D. Walton (2006). The Carneades argumentation framework - using presumptions and exceptions to model critical questions. In P. Dunne and T. Bench-Capon (Eds.), *Computational Models of Argument. Proceedings of COMMA-06*, pp.195–207. IOS Press.

Gordon, T. F. and D. N. Walton (2009). Legal reasoning with argumentation schemes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law (ICAIL 2009)*. ACM.

Hart, H. L. A. (1994). Postscript. In The Concept of Law, pages 238–76. Oxford University Press, 2nd edition.

Horty, J. F. (2012). Reasons as Defaults. Oxford University Press.

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

Loui, R. P. and Norman, J. (1995). Rationales and argument moves. Artificial Intelligence and Law, 3:159–89.

Macagno, F., Walton, D., and Sartor, G. (2018). Pragmatic maxims and presumptions in legal interpretation. Law and Philosophy, 37:69–115.

MacCormick, N. (2005). *Rhetoric and the Rule of Law: A Theory of Legal Reasoning*. Oxford: Oxford University Press.

Maranhao, J. d. S. and G. Sartor (2019). Value assessment and revision in legal interpretation. In *ICAIl 2019*, 219–223. ACM.

Marmor, A. and Soames, S. (2011). Language in the Law. Oxford University Press.

Modgil, S. and H. Prakken (2013). A general account of argumentation with preferences. *Artificial Intelligence 195*, 361–97.

Patterson, D. (2005). Interpretation in law. San Diego Law Review, 42.

Pollock, J. L. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT.

Pollock, J. L. (2008). Defeasible reasoning. In J. E. Adler and L. J. Rips (Eds.), *Reasoning: Studies of Human Inference and its Foundations*, pp. 451–70. Cambridge University Press.

Prakken, H. (2010). On the nature of inference schemes. In C. Reed and C. Tindale (Eds.), *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument*, pp. 167–185. College Publications.

Prakken, H. (2019). Modelling accrual of arguments in ASPIC+. In ICAIL 2019, 103–112. ACM.

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

Prakken, H. and G. Sartor (2006). Presumptions and burdens of proof. In Van Engers, T., editor, Proceedings of the Nineteenth Annual Conference on Legal Knowledge and Information Systems (JURIX), pages 176–185. IOS.

Bench-Capon, T. J. M. and Rissland, E. L. (2002). A note on dimensions and factors. *Artificial Intelligence and Law*, 10_ 65–77.

Riveret, R., P. Baroni, Y. Gao, G. Governatori, A. Rotolo, and G. Sartor (2018). A labelling approach to probabilistic argumentation. *Annals of Mathematics and Artificial Intelligence*, 83:21–71.

Rotolo, A., G. Governatori, and G. Sartor (2015). Deontic defeasible reasoning in legal interpretation: Two options for modelling interpretive arguments. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law (ICAIL'05)*, 99–108. ACM.

Sartor, G. (2018). Defeasibility in law. In Bongiovanni, G., Postema, G., Rotolo, A., Sartor, G., Valentini, C., and Walton, D., editors, Handbook of Legal Reasoning and Argumentation, 315–64. Springer.

Scalia, A. and B. A. Garner (2012). *Reading law. The interpretation of legal texts*. Thomson West.

Soames, S. (2013). Deferentialism: A post–originalist theory of legal interpretation. Fordham Law Review, 82:597–613.

Solum, L. B. (2010). The interpretation-construction distinction. Constitutional Commentary, 27:95–118.

Toulmin, S. ([1958] 2003). *The Uses of Argument*. Cambridge University Press.

Walton, D., G. Sartor, and F. Macagno (2018). Statutory interpretation as argumentation. In Bongiovanni, G., Postema, G., Rotolo, A., Sartor, G., Valentini, C., and Walton, D., editors, *Handbook of Legal Reasoning and Argumentation*, 519–60. Springer.

Walton, D., Macagno, F., and Sartor, G. (2020). *The Pragmatic and Argumentative Structure of Statutory Interpretation.* Cambridge University Press (forthcoming).

Zitelmann, E. (1903). *Lücken im Recht*. Duncker-Humblot.