Formalising debates about law-making proposals as practical reasoning

Henry Prakken

Department of Information and Computing Sciences, Utrecht University, and

Faculty of Law, University of Groningen

May 26, 2014

Abstract

In this paper the $ASPIC^+$ framework for argumentation-based inference is used for formally reconstructing two legal debates about law-making proposals: an opinion of a legal scholar on a Dutch legislative proposal and a US common-law judicial decision on whether an existing common law rule should be followed or distinguished. Both debates are formalised as practical reasoning, with versions of the argument schemes from good and bad consequences. These case studies aim to contribute to an understanding of the logical structure of debates about law-making proposals. Another aim of the case studies is to provide new benchmark examples for comparing alternative formal frameworks for modelling argumentation. In particular, this paper aims to illustrate the usefulness of two features of $ASPIC^+$: its distinction between deductive and defeasible inference rules and its ability to express arbitrary preference orderings on arguments.

Keywords: Law making debates, practical reasoning, argumentation, formalisation, argument schemes.

1   Introduction

Modern approaches to legal logic account for the fact that legal reasoning is not only about *constructing* arguments but also about *attacking* and *comparing* them. This is partly since legal reasoning often takes place in adversarial contexts (the court room, parliament). But even an individual legal reasoner (judge, solicitor, politician or politically interested citizen) often considers reasons for and against claims or proposals. Modern logic provides tools for formalising such argumentative reasoning. This paper [1] aims to provide an illustration of the usefulness of these tools, in the form of two case studies of how law-making debates can be formalised in an argumentation logic. In the first case study an opinion of a legal scholar on a Dutch legislative proposal is formalised, while in the second case study a judicial decision in the US common law of contract is reconstructed. Both case studies employ the $ASPIC^+$ framework for argumentation (Prakken, 2010; Modgil and Prakken, 2013), which currently is one of the main logical frameworks for argumentation in the field of artificial intelligence (AI). The $ASPIC^+$ framework has been applied earlier in a realistic case study in Prakken

---

[1]This paper is an extended and revised version of Prakken (2012a). The use of recursive labellings in $ASPIC^+$ is new, Section 5 is new, and the text of the other sections has been extended.

(2012b); in that paper the main arguments were not about law making proposals but about interpreting and applying legal concepts.

Both case studies concern law-making debates, one about a proposal for legislation in a civil law jurisdiction and the other in the context of common law precedent. While thus the legal context is different in the case studies, it will turn out that the reasoning forms are quite similar and are instances of what philosophers call practical reasoning, that is, reasoning about what to do. In particular, in both cases use is made of so-called argument schemes of good and bad consequences of decisions for action. Recently, these schemes have received much attention in the AI (& Law) literature. In this paper they will be formalised as proposed in Bench-Capon and Prakken (2010); Bench-Capon et al. (2011). Unlike other formulations of these schemes, these formulations do not refer to *single* but to *sets* of consequences of actions, thus allowing for aggregation of reasons for and against proposals. The present paper's main advance over Bench-Capon and Prakken (2010); Bench-Capon et al. (2011) is that it models an actual example of a legal argument in its full detail instead of modelling a simplified example that is more loosely based on actual textual material.

Another aim of the two case studies in this paper is to provide new benchmark examples for comparing alternative formal frameworks for modelling argumentation. In both general AI and AI & law several formal frameworks for argumentation-based inference have been proposed, such as assumption-based argumentation (Bondarenko et al., 1997), classical argumentation (Besnard and Hunter, 2008), Carneades (Gordon et al., 2007) and $ASPIC^+$. This raises the question which framework is best suited for formalising natural, in particular legal arguments. The present paper aims to contribute to this discussion. While case studies cannot decide which framework is the best, they help in providing evidence and formulating benchmark examples. Compared to assumption-based and classical argumentation, the main distinguishing features of $ASPIC^+$ are an explicit distinction between deductive and defeasible inference rules and an explicit preference ordering on arguments. Accordingly, one aim of the present case studies is to illustrate the usefulness of these features.

This paper is organised as follows. First in Section 2 the idea of logical argumentation systems is introduced, after which in Section 3 the $ASPIC^+$ framework is reviewed. Then in Section 4 the Dutch legal opinion is presented, which is reconstructed in $ASPIC^+$ in Section 5. In Section 5 the Monge case from US common contract law is presented and formalised. The paper concludes in Section 7.

2 Introduction to logical argumentation systems

Logical research in AI & Law has recognised from the start that legal reasoning is defeasible and that therefore some form of nonmonotonic logic is needed to formalise legal argument, that is, a logic that allows that valid conclusions can be invalidated by further information. While in the early days of AI &Law nonmonotonic logic of several kinds were used, such as Reason-Based Logic of Hage (1997) and Verheij (1996), nowadays argumentation-based logics are the most commonly used. Such systems formalise defeasible reasoning as the construction and comparison of arguments for and against certain conclusions. An argument only warrants its conclusion if firstly, it is properly constructed and, secondly, it can be defended against counterarguments. Thus argumentation logics define three things: how arguments can be constructed, how they can be attacked by counterarguments and how they can be defended against such attacks. In general, three kinds of *attack* are distinguished: arguing for a contradictory

conclusion, arguing that an inference rule has an exception, or denying a premise. An argument A is then said to *defeat* an argument B if A attacks B and is not weaker than B. The relative strength between arguments is determined with any standard that is appropriate to the problem at hand and may itself be the subject of argumentation. Note that if two arguments attack each other and are equally strong or their relative strength cannot be determined, then they defeat each other. The defeasibility of arguments arises from the fact that new information may give rise to new counterarguments that defeat the original argument.

To determine which arguments are acceptable, it does not suffice to determine the defeat relations between two arguments that attack each other. We must also look at how arguments can be *defended* by other arguments. Suppose we have three arguments A, B and C such that B strictly defeats A and C strictly defeats B. Then C defends A against B so, since C is not attacked by any argument, both A and C (and their conclusions) are acceptable while B is not acceptable. However, we can easily imagine more complex examples where our intuitions fall short. For instance, another argument D could be constructed such that C and D defeat each other, then an argument E could be constructed that defeats D but is defeated by A, and so on: which arguments can now be accepted and which should be rejected? Here we cannot rely on intuitions but need a precise formal definition. Such a definition should dialectically assess all constructible arguments in terms of three classes (three and not two since some conflicts cannot be resolved). Intuitively, the *justified arguments* are those that survive all conflicts with their attackers and so can be accepted, the *overruled arguments* are those that are defeated by a justified argument and so must be rejected; and the *defensible arguments* are those that are involved in conflicts that cannot be resolved. Furthermore, a *statement is justified* if it has a justified argument, it is *overruled* if all arguments for it are overruled, and it is *defensible* if it has a defensible argument but no justified arguments. In terms more familiar to lawyers, if a claim is justified, then a rational adjudicator is convinced that the claim is true, if it is overruled, such an adjudicator is convinced that the claim is false, while if it is defensible, s/he is neither convinced that it is true nor that it is false.

3   The *ASPIC*$^+$ framework

In this section we review the *ASPIC*$^+$ framework of Prakken (2010) and Modgil and Prakken (2013). It defines arguments as inference trees formed by applying strict or defeasible inference rules to premises formulated in some logical language. Informally, if an inference rule's antecedents are accepted, then if the rule is strict, its consequent must be accepted *no matter what*, while if the rule is defeasible, its consequent must be accepted *if there are no good reasons not to accept it*. Arguments can be attacked on their (non-axiom) premises and on their applications of defeasible inference rules. Some attacks succeed as *defeats*, which is partly determined by preferences. The acceptability status of arguments is then defined by checking whether an argument can be defended against all its defeaters.

ASPIC$^+$ is not a system but a framework for specifying systems. It defines the notion of an abstract *argumentation system* as a structure consisting of a logical language $\mathcal{L}$ closed under negation, a set $\mathcal{R}$ consisting of two subsets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules, and a naming convention n in $\mathcal{L}$ for defeasible rules in order to talk about the applicability of defeasible rules in $\mathcal{L}$. Thus, informally, $n(r)$ is a well-formed formula in $\mathcal{L}$ which says that rule $r \in \mathcal{R}$ is applicable. (as is usual,

the inference rules in $\mathcal{R}$ are defined *over* the language $\mathcal{L}$ and are not elements *in* the language.)

$ASPIC^+$ does not commit to a particular logical language or to particular sets of inference rules. For $\mathcal{L}$ any logical language can be chosen, such as the language of propositional logic, first-order predicate logic or deontic logic. $ASPIC^+$'s inference rules can be used in two ways: they could encode domain-specific information (such as commonsense generalisations or legal rules) but they could also express general laws of reasoning. When used in the latter way, the strict rules over $\mathcal{L}$ can be based on the semantic interpretation of $\mathcal{L}$ by saying that $\mathcal{R}_s$ contains all inference rules that are semantically valid over $\mathcal{L}$ (according to the chosen semantics). So, for example, if $\mathcal{L}$ is chosen to be the language of standard propositional logic, then $\mathcal{R}_s$ can be chosen to consist of all semantically valid inferences in standard propositional logic (whether such an inference is valid can be tested with, for example, the truth-table method).

The defeasible inference rules $\mathcal{R}_d$ cannot be based on the semantic interpretation of $\mathcal{L}$, since they go beyond the meaning of the logical constants in $\mathcal{L}$. Consider, for example, defeasible modus ponens: 'if P then usually Q' and P do not together deductively imply Q, since we could have an unusual case of P. In other words, defeasible inference rules are deductively invalid. They can instead be based on insights from epistemology or argumentation theory. For example, $\mathcal{R}_d$ could be filled with presumptive argument schemes in the sense of Walton (1996) and Walton et al. (2008). The critical questions of these schemes are then pointers to counterarguments.

In $ASPIC^+$ argumentation systems are applied to knowledge bases to generate arguments and counterarguments. Combining these with an argument ordering results in so-called argumentation theories.

Definition 1 [Argumentation systems] An *argumentation system* is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language closed under negation ($\neg$).

- $\mathcal{R}_s$ and $\mathcal{R}_d$ are two disjoint sets of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over well-formed formulas in $\mathcal{L}$).

- n is a naming convention for defeasible rules, which to each rule r in $\mathcal{R}_d$ assigns a well-formed formula φ from $\mathcal{L}$ (written as $n(r) = \varphi$).

We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$.

Definition 2 [Knowledge bases] A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the *axioms*) and $\mathcal{K}_p$ (the *ordinary premises*).

Intuitively, the axioms are certain knowledge and thus cannot be attacked, whereas the ordinary premises are uncertain and thus can be attacked.

Arguments can be constructed step-by-step from knowledge bases by chaining inference rules into trees. Arguments thus contain subarguments, which are the structures that support intermediate conclusions (plus the argument itself and its premises as limiting cases). In what follows, for a given argument A the function `Prem` returns all its premises, `Conc` returns its conclusion, `TopRule` returns the final rule applied in the argument, `Sub` returns all its sub-arguments and `ImmSub` returns all its immediate sub-arguments, i.e., the subarguments to which' conclusions the argument's top rule was applied.

Definition 3 [**Arguments**] An *argument* A on the basis of a knowledge base KB in an argumentation system $(\mathcal{L}, \mathcal{R}, n)$ is:

1. φ if $\varphi \in \mathcal{K}$ with:
   $\text{Prem}(A) = \{\varphi\}$;
   $\text{Conc}(A) = \varphi$;
   $\text{TopRule}(A) = $ undefined;
   $\text{Sub}(A) = \{\varphi\}$;
   $\text{ImmSub}(A) = \emptyset$.

2. $A_1, \ldots A_n \rightarrow/\Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$, with
   $\text{Prem}(A) = \text{Prem}(A_1) \cup \ldots \cup \text{Prem}(A_n)$;
   $\text{Conc}(A) = \psi$;
   $\text{TopRule}(A) = \text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow/\Rightarrow \psi$;
   $\text{Sub}(A) = \text{Sub}(A_1) \cup \ldots \cup \text{Sub}(A_n) \cup \{A\}$;
   $\text{ImmSub}(A) = \{A_1, \ldots, A_n\}$.

Example 1 Consider a knowledge base in an argumentation system with

$$\mathcal{R}_s = \{p, q \rightarrow s; \ u, v \rightarrow w\}; \mathcal{R}_d = \{p \Rightarrow t; \ s, r, t \Rightarrow v\}$$
$$\mathcal{K}_n = \{q\}; \mathcal{K}_p = \{p, u, r\}$$

An argument for w is displayed in Figure 1. The type of a premise is indicated with a superscript and defeasible inferences and attackable premises and conclusions are displayed with dotted lines.
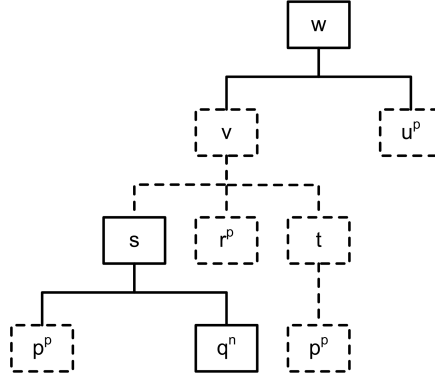

Figure 1: An argument

Formally the argument and its subarguments are written as follows:

$A_1: p$ $A_5: A_1 \Rightarrow t$
$A_2: q$ $A_6: A_1, A_2 \rightarrow s$
$A_3: r$ $A_7: A_5, A_3, A_6 \Rightarrow v$
$A_4: u$ $A_8: A_7, A_4 \rightarrow w$

We have that

5

$$
\begin{aligned}
\texttt{Prem}(A_8) &= \{p, q, r, u\} \\
\texttt{Conc}(A_8) &= w \\
\texttt{Sub}(A_8) &= \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\} \\
\texttt{ImmSub}(A_8) &= \{A_4, A_7\} \\
\texttt{DefRules}(A_8) &= \{p \Rightarrow t;\ s, r, t \Rightarrow v\} \\
\texttt{TopRule}(A_8) &= u, v \rightarrow w
\end{aligned}
$$

Arguments can be attacked in three ways: on their premises (undermining attack), on their conclusion (rebutting attack) or on an inference step (undercutting attack). The latter two are only possible on applications of defeasible inference rules.

Definition 4 [**Attack**] *A attacks B* iff *A undercuts*, *rebuts* or *undermines B*, where:

- *A undercuts* argument B (on B') iff $\texttt{Conc}(A) = -n(r)$ for some $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.

- *A rebuts* argument B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.

- Argument *A undermines* B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' = \varphi$, $\varphi \notin \mathcal{K}_n$.

The argument in Example 1 can be undermined on any premise except on q, it can be rebutted by arguments with a conclusion $\neg t$ or $\neg v$ and it can be undercut by arguments with a conclusion $\neg r_1$ and $\neg r_2$, assuming that $n(p \Rightarrow t) = r_1$ and $n(s, r, t \Rightarrow v) = r_2$.

Argumentation systems plus knowledge bases form argumentation theories, which induce structured argumentation frameworks.

Definition 5 [**Structured Argumentation Frameworks**] Let AT be an *argumentation theory* (AS, KB). A *structured argumentation framework (*SAF*)* defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all finite arguments constructed from KB in AS, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

The notion of *defeat* can then be defined by using the argument ordering to check which attacks succeed as defeats. Assumptions could be made on the properties of $\preceq$ (such as that it is transitive) but but the definition of defeat does not rely on any assumption. In fact, undercutting attacks succeed as *defeats* independently of preferences over arguments, since they express exceptions to defeasible inference rules. By contrast, rebutting and undermining attacks succeed only if the attacked argument is not stronger than the attacking argument. ($A \prec B$ is defined as usual as $A \preceq B$ and $B \not\preceq A$).

Definition 6 [**Defeat**] *A defeats B* iff:

- *A* undercuts B; or

- *A* rebuts/undermines B on B' and $A \not\prec B'$.

A strictly *defeats B* iff *A defeats B* and *B* does not *defeat A*

The success of rebutting and undermining attacks thus involves comparing the conflicting arguments at the points where they conflict. The definition of successful undermining exploits the fact that an argument premise is also a subargument.

The final task is to define how the arguments of an argumentation theory can be evaluated in the context of all arguments in the theory and their defeat relations. The following definition of recursive argument labellings, originally proposed by Pollock (1995), achieves this.[2] It uses the notion of an *immediate* subargument of an argument. This notion was in Definition 3 defined as $\texttt{ImmSub}(A)$, that is, as those arguments that provide the antecedents of the top rule of argument A. Note that arguments taken from $\mathcal{K}$ thus have no immediate subarguments.

The definition of recursive argument labellings uses the notion of direct defeat. That an argument A *directly defeats* an argument B means that A rebuts, undercuts or undermines B on B (and $A \not\prec B$ in case A rebuts or undermines B).

Definition 7 [Recursive argument labellings] For any structured argumentation framework $SAF = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$, a *p-labelling* of SAF is a pair of sets (In, Out) (where both In and Out are subsets of $\mathcal{A}$) such that $In \cap Out = \emptyset$ and for all arguments A in $\mathcal{A}$ it holds that:

1. argument A is labelled *in* iff:

    (a) all arguments in $\mathcal{A}$ that directly defeat A are labelled *out*; and

    (b) all immediate subarguments of A are labelled *in*; and

2. argument A is labelled *out* iff:

    (a) A is directly defeated by an argument in $\mathcal{A}$ that is labelled *in*; or

    (b) An immediate subargument of A is labelled *out*.

This definition implies that an argument is *out* if at least one of its subarguments is *out*. Note also that according to this definition not all arguments have to be labelled. For example, if the argumentation theory contains just two arguments A and B, which defeat each other, then $(\emptyset, \emptyset)$ is a well-defined labelling. Moreover, in general the set of all arguments can be labelled in more than one way that satisfies this definition. For instance, in our example two further well-defined labellings are respectively, a labelling in which A is *in* while B is *out* and a labelling in which B is *in* while A is *out*. To further select from these well-defined labellings, several labelling policies are possible, which correspond to different so-called semantics for argument evaluation (cf. Caminada (2006)). We discuss two of them. *Grounded semantics* minimises the set of all arguments that are labelled *in*.). So in our example, only $(\emptyset, \emptyset)$ is a grounded labelling. *Preferred semantics* instead maximises the set of arguments that are labelled *in*. So in our example the two labellings that label one argument *in* and the other *out* are the two preferred labellings. It is known that the grounded labelling is always unique (since if an argument can both be labelled *in* and labelled *out*, it leaves the argument unlabelled), while preferred semantics allows for alternative labellings (since if an argument can both be labelled *in* and labelled *out*, it alternatively explores both choices). In this paper preferred semantics will be used, since it allows for identifying alternative coherent positions.

Finally, in preferred semantics an argument is *justified* if it is labelled *in* in all labellings, it is *overruled* if it is labelled *out* in all labellings, and it is *defensible* if it is

---

[2]In previous publications on *ASPIC$^+$* arguments were instead evaluated by generating a so-called abstract argumentation framework from an argumentation theory and evaluating arguments with any of the abstract semantics of Dung (1995). While this is theoretically fine, in Prakken (2013) I argued that Pollock (1995)'s recursive labellings support a more natural explanation of argument evaluation. I also proved that the two ways to evaluate arguments always yield the same outcome, so that logically their differences do not matter.

neither justified nor overruled. Furthermore, a statement is *justified* if it is the conclusion of a justified argument, while it is *defensible* if it is not justified but the conclusion of a defensible argument, and *overruled* if it is defeated by a justified argument.

4 An example of natural argument

The following text is a summary of an opinion by Nico Kwakman of the Faculty of Law, University of Groningen, The Netherlands.[3] The topic is whether the legislative proposal by the Dutch government to impose mandatory minimum sentences for serious crimes is a good idea.

> Despite strong criticism from the Council of State (Raad van State, RvS), the Cabinet is going to continue to introduce mandatory minimum sentences for serious offences. Dr Nico Kwakman, criminal justice expert at the University of Groningen, is critical of the bill, but can also understand the reasoning behind it. The effectiveness of the bill is doubtful, but the symbolic impact is large. The cabinet is sending out a strong signal and it has every right to do so.
>
> The Netherlands Bar Association, the Council of State, the Netherlands Association for the Judiciary, they are all advising the cabinet not to introduce the bill. However, the cabinet is ignoring their advice and continuing on with its plans. Criminals who commit a serious crime for the second time within ten years must be given a minimum sentence of at least half of the maximum sentence allocated to that offence, says the Cabinet. The bill has been drawn up under great pressure from the PVV party.
>
> **Not effective**   Regarding content, the bill raises a lot of question marks, explains Kwakman. Heavy sentences do not reduce the chances of recidivism, academic research has revealed. Nor has it ever been demonstrated that heavy sentences lead to a reduction in the crime figures. Kwakman: 'It is very important for a judge to be able to tailor a punishment to the individual offender. That increases the chances of a successful return to society. In the future, judges will have much less room for such tailoring.'
>
> **Call from the public**   The Cabinet says that the new bill is meeting the call from the public for heavier sentences. This is despite the fact that international comparisons show that crime in the Netherlands is already heavily punished. Kwakman: 'Dutch judges are definitely not softies, as is often claimed. Even without politics ordering them to, in the past few years they have become much stricter in reaction to what is going on in society. This bill, completely unnecessarily, will force them to go even further'.
>
> **Symbolic impact**   Kwakman does have a certain amount of sympathy for the Cabinet's reasoning. 'The effectiveness of the bill is doubtful, but criminal law revolves around more than effectiveness alone. It will also have a significant symbolic impact. The Cabinet is probably mainly interested in the symbolism, in underlining norms. The Cabinet is sending out a strong signal and it has every right to do so as the democratically elected legislator. Anyone who doesn't agree should vote for a different party the next time.'

---

[3]Published at http://www.rug.nl/news-and-events/people-perspectives/opinie/2012/06nicokwakman?lang=en on on 29 February 2012.

French kissing is rape   Judges currently have a lot of freedom when setting sentences but that will be significantly less in the future. Kwakman: 'A forced French kiss is a graphic example. It officially counts as rape, but judges impose relatively mild sentences for it. Soon judges will be forced to impose half of the maximum punishment for rape on someone who is guilty of a forced French kiss for the second time. Only in extremely exceptional cases can that sentence be changed.'

Taking a stand   And that is where the dangers of the new bill lurk, thinks Kwakman. Judges who don't think the mandatory sentence is suitable will look for ways to get around the bill. These could include not assuming so quickly that punishable offences have been proven, interpreting the bill in a very wide way on their own initiative, or by thinking up emergency constructions. Kwakman: 'In this way judges will be taking on more and more of the legislative and law formation tasks, and that is a real shame. The legislature and the judiciary should complement each other. This bill will force people to take a stand and the relationship between legislator and judge will harden.'

5   A formal reconstruction in *ASPIC*$^+$

I next model the example of the previous section in the *ASPIC*$^+$ framework, leaving the logical language formally undefined and instead using streamlined natural language for expressing the premises and conclusions of the arguments. Argument schemes are modelled as defeasible inference rules. The case is reconstructed in terms of argument schemes from good and bad consequences recently proposed by Bench-Capon et al. (2011) and some other schemes. Contrary to the usual formulations of schemes from consequences (e.g. Walton et al. (2008); Atkinson and Bench-Capon (2007)), they do not refer to single but to *sets of* good or bad consequences.[4] Thus argumentation can be modelled as collecting and then weighing all good and bad consequences of alternative action proposals. An early application of this idea in Reason-Based logic was proposed by Hage (2004). Current work generally respects Hage's insights but formalises them in the context of an argumentation logic.

Argument scheme from good consequences

Action A results in C_1

...

Action A results in C_n

C_1 is good

...

C_n is good

Therefore (presumably), action A is good.

Argument scheme from bad consequences

---

[4]As usual, inference rules with free variables are schemes for all their ground instances.

Action A results in C_1

...

Action A results in C_m

C_1 is bad

...

C_m is bad

_____

Therefore (presumably), action A is bad.

These schemes have four critical questions:

1. Does A result in $C_1, \ldots, C_n/C_m$?
2. Is $C_1, \ldots, C_n/C_m$ really good/bad?
3. Does A also result in something which is bad (good)?
4. Is there another way to realise C_n/C_m?

In *ASPIC*$^+$ these questions are pointers to counterarguments. Questions 1 and point to underminers, question 3 to rebuttals and question 4 to undercutters. Note that if there is more than one good (bad) consequence of a given action, then the scheme of good (bad) consequences can be instantiated several times, namely for each combination of one or more of these consequences. This makes it possible to model a kind of accrual, or aggregation of reasons for or against an action proposal.

My reconstruction of Kwakman's opinion is visualised in Figure 2. In this figure, solid lines stand for applications of inference rules (with their antecedents below and their consequent above). A solid line that branches out toward below indicates an inference rule applied to multiple antecedents. The three dotted lines indicate direct attack relations. The four boxes with thick borders are the 'final' conclusions of the four largest arguments. Finally, the grey colourings of some nodes will be explained later.

All arguments in my reconstruction either instantiate one of these schemes or attack one of their premises, using another argument scheme, which I now informally specify: (all inferences in Figure 2 are labelled with the name of the inference rule that they apply):

- GCi and BCi stand for, respectively, the i'th application of the scheme from good, respectively, bad consequences.

- D stands for the application of a definition in a deductive inference:

  P (categorically/presumably) causes Q

  Q is by definition a case of R

  _____

  Therefore (strictly), P (categorically/presumably) causes R

- C1 and C2 stand for two applications of causal chaining:

  P_1 (categorically/presumably) causes P_2

  P_2 (categorically/presumably) causes ...

  ...(categorically/presumably) causes P_n

  _____

  Therefore (strictly/presumably), P_1 causes P_n

  This inference rule is strict or defeasible depending on whether the causal relations are assumed to be categorical or presumptive.

- DMP stands for defeasible modus ponens:

  If P_1 and ... and P_n then usually/typically/normally Q

  P_1 and ... and P_n

  Therefore (presumably) Q

- SE is shorthand for a 'scientific evidence' scheme:

  Scientific evidence shows that P

  Therefore (presumably) P

The links in Figure 2 to the final two conclusions require some explanation. If there is a set S of reasons why action A is good, then the scheme from good consequences can be instantiated for any nonempty subset of S. This is informally visualised by introducing a name on the support links for any of these reasons. This summarises all possible instances of the scheme from good consequences. Thus in the example there are seven such instances, one combining GC1, GC2 and GC3 (denoted below by GC_{123}), three with any combination of two reasons (denoted below by $GC_{12}, GC_{13}, GC_{23}$) and three applying any individual reason (denoted below by G_1, G_2 and G_3). Likewise, there are three instances of the scheme from bad consequences, two applying an individual reason for a bad consequence (BC_1 and BC_2) and one combining these reasons (BC_{12}). Below we will see that this complicates the identification of the various preferred labellings.

The argumentation system and knowledge base corresponding to Figure 2 can be summarised as follows:

- $\mathcal{L}$ is a first-order predicate-logic language (here informally presented), where for ease of notation 'Action A is good' and 'Action A is bad' are regarded as negating each other.

- R_s contains at least the D rule mentioned above, and it contains the C rule if the causal relations in the example to which it is applied are regarded as categorical. Furthermore, it contains all deductively valid propositional and first-order predicate-logic inferences.

- R_d consists of the argument schemes from good and bad consequences, the C rule if not included in $\mathcal{R}_s$, and the SE and DMP rules.

- $\mathcal{K}_n$ is empty, while $\mathcal{K}_p$ consists of the leafs of the four argument trees (where their conclusions are regarded as their roots). $\mathcal{K}$ thus consists of 18 ordinary premises.

The argumentation theory induced by this argumentation system and this knowledge base is as follows:

- $\mathcal{A}$ consists of quite a number of arguments:

  - all 18 premises;
  - two applications of the C rule: C_1 and C_2;
  - one application of the DMP rule: DMP;
  - one application of the D rule: D;
  - seven applications of the GC scheme: $GC_1, GC_2, GC_3, GC_{12}, GC_{13}, GC_{23}, GC_{123}$;

– three applications of the BC scheme: BC_1, BC_2, BC_{12}.

So in total the reconstruction contains 29 arguments. Note that all 11 non-premise arguments contain other arguments from $\mathcal{A}$ as their subarguments.

- The attack relations are more in number than the three shown in Figure 2:

    – Any argument applying GC rebuts any argument applying BC and vice versa;

    – C_1 undermines the premise argument P_1 = 'The act will reduce recidivism' and all arguments using it, that is, the arguments D, GC_1, GC_{12}, GC_{13}, GC_{123};

    – The premise argument P_1 in turn rebuts argument C_1;

    – DMP undermines the premise argument P_2 = 'Meeting the call for the public for heavier sentences is good' and all arguments using it, that is, GC_2, GC_{12}, GC_{23}, GC_{123};

    – The premise argument P_2 in turn rebuts argument DMP.

- Various argument orderings can be assumed, resulting in different defeat relations. Note that the argument ordering is only applied to 'direct' attacks, namely, to the attacks between C_1 and P_1, between C_2 and P_2, and between all applications of the GC scheme and all applications of the BC scheme.

Let us now for simplicity assume that the argument ordering counts reasons for and against an action, and moreover that, for whatever reason, $P_1 \prec C_1$ while $DMP \approx P_2$.[5]

What are now the preferred labellings? To determine them, we must take into account that Figure 2 in fact summarises seven applications of the scheme from good consequences and three applications of the scheme from bad consequences. So strictly speaking the conclusion that passing the act is good should be multiplied seven times in Figure 2 and the conclusion that passing the act is bad should be tripled. This would clutter the graph and make it poorly understandable. Fortunately, we can simplify our analysis as follows. Note first that GC_1 is always *out* since its subargument P_1 is directly defeated by C_1, which has no defeaters and is therefore always *in*. SO P_1 is always *out*. But then D is always *out* since it has an immediate subargument that is *out* and so for the same reason GC_1 is always *out*. By the same line of reasoning GC_{12}, GC_{13} and GC_{123} are also always *out* since they have a subargument (P_1) that is always *out*. Furthermore, note that argument GC_{23} is stronger in the argument ordering than both GC_2 and GC_3, since the argument ordering counts the number of good and bad consequences. Moreover, GC_{23} has no attackers that do not also attack either GC_2 or GC_3, so we can safely ignore GC_2 and GC_3. We can therefore safely assume in Figure 2 that the statement that passing the act is good is the conclusion of GC_{23}. For similar reasons we can safely assume in Figure 2 that the statement that passing the act is bad is the conclusion of BC_{12}.

Now there are two conflicts between equally strong arguments in Figure 2 that induce alternative preferred labellings (recall that if an argument can be both labelled *in* and labelled *out*, preferred semantics always explores both options). Consider first the conflict between DMP and P_2. We can make DMP *in* if we make P_2 *out*, since

all subarguments of DMP are *in* since they have no defeater. But then GC_{23} has a subargument that is *out* so GC_{23} is also *out*. Then BC_{12} is *in* since, firstly, its only defeater is *out* and, second, all its subarguments are *in* since none of them has a defeater. The resulting labelling is displayed in Figure 2, in which grey boxes are conclusions of arguments that are *out* while white boxes are conclusions of arguments that are *in*(so in this labelling there are no unlabelled arguments).

Alternatively, we can make P_1 *in* and DMP *out*. Then we have to consider the conflict between GC_{23} and BC_{12}. For both of them it now holds that all their subarguments are *in*. So we have two options: make GC_{23} *in* and BC_{12} *out* or vice versa. For reasons of space we display only the first of these labellings, in Figure 3. The alternative labelling can be visualised by just switching the labels of GC_{23} and BC_{12}.

In sum, there are both labellings where GC_{23} is *in* and BC_{12} is *out* and labellings where GC_{23} is *out* and BC_{12} is *in*. Therefore, both the conclusion that passing the act is good and the conclusion that passing it is bad are defensible. To make the conclusion that passing the act is good justified, one should either argue that DMP is strictly preferred over P_2 or argue that for some reason the two good consequences 2 and 3 together outweigh the two bad consequences 1 and 2.

6 Law making debates in case law: the Olga Monge case

Above I illustrated how legislative debates can be reconstructed as practical reasoning. In this section I illustrate that the same is sometimes possible for common-law judicial decisions about whether to follow or to distinguish a common-law rule. I illustrate this with an American common law of contract case, the *Olga Monge v. Beebe Rubber Company* case, decided by the Supreme Court of New Hampshire (USA), February 28, 1974. In brief, the facts were that Olga Monge, according to the court "a virtuous mother of three", was employed at will (that is, for an indefinite period of time) by Beebe Rubber Company. The relevant common law rule at that time said that every employment contract that specifies no duration is terminable at will by either party, which means that the employee can be fired for any reason or no reason at all. At some point, Olga Monge was fired for no reason by her foreman. Olga claimed that this was since she had refused to go out with him and she claimed breach of contract, arguing that the common law rule does not apply if the employee was fired in bad faith, malice, or retaliation. The court accepted that she was fired was that reason and was then faced with the problem whether to follow the old rule and decide that there was no breach of contract, or to distinguish the rule into a new rule by adding an exception in case the employee was fired in bad faith, malice, or retaliation, in order to decide that there was breach of contract. Here it is relevant that according to one common law theory of precedential constraint, courts can distinguish an old rule by adding an extra condition as long as the new rule still gives the same outcome in all precedent cases as the old rule. See Horty (2011); Horty and Bench-Capon (2012) for a discussion and formalisation of this theory.

The court decided to distinguish the old rule, on the following grounds:

> In all employment contracts, whether at will or for a definite term, the employer's interest in running his business as he sees fit must be balanced against the interest of the employee in maintaining his employment, and the public's interest in maintaining a proper balance between the two.

(...)

> We hold that a termination by the employer of a contract of employment at will which is motivated by bad faith or malice or based on retaliation is not in the best interest of the economic system or the public good and constitutes a breach of the employment contract.

I now reconstruct this reasoning as practical reasoning with the argument scheme from good consequences. The two alternative decisions are to follow the old rule or to distinguish it into the new rule by adding a condition 'unless the employee was fired in bad faith, malice, or retaliation'. In my interpretation the court stated as a good consequence of following the old rule that the employer's interest in running his business as he sees fit are protected while it stated a good consequence of distinguishing it promotes the interest of the economic system and the public good. We then have two instances of the argument scheme from good consequences for conflicting decisions. The conclusion of both of these arguments is then combined with an argument that applies the adopted rule. The resulting reconstruction is visualised in Figure 4. For space limitations we leave implicit that if Olga Monge could (could not) be fired for no reason, then firing her for no reason was not (was) breach of contract.

Two arguments in this reconstruction apply the argument scheme from a single good consequence GC1. One argument applies the causal chaining scheme C. Two arguments apply the classical modus ponens inference rule MP. Finally, the two top rules of the rebutting arguments for whether Olga Monge could be fired for no reason apply defeasible modus ponens on the Old, respectively, the new Rule (where the second application of defeasible modus ponens is in fact applied to the 'only if' part of the new Rule).

The *ASPIC*$^+$ argumentation system and knowledge base corresponding to Figure 4 can be summarised as follows:

- $\mathcal{L}$ is as above a first-order predicate-logic language (here informally presented), where for ease of notation 'We should adopt the Old Rule as the valid rule' and 'We should adopt the New Rule as the valid rule' are regarded as negating each other. Furthermore, we assume that $\mathcal{L}$ has a defeasible connective for representing legal rules.

- R_s contains all deductively valid propositional and first-order predicate-logic inferences.

- R_d consists of defeasible modus ponens for legal rules, the two argument schemes from good and bad consequences and the C rule.

- $\mathcal{K}_n$ is empty, while $\mathcal{K}_p$ consists of the leafs of the two argument trees (where their conclusions are regarded as their roots). $\mathcal{K}$ thus consists of 8 ordinary premises.

The *ASPIC*$^+$ argumentation theory induced by this argumentation theory is as follows:

- $\mathcal{A}$ consists of the following arguments:

    - all 8 premises;
    - one application of the C rule: C;
    - two applications of the modus ponens rule: MP_1 and MP_2;
    - two applications of the GC scheme: GC_{1a} and GC_{1b};

- two applications of defeasible modus ponens on legal rules: DMP_1 and DMP_2.

So in total the reconstruction contains 15 arguments.

- The attack relations are again more in number than the two shown in Figure 4:
  - Arguments DMP_1 and DMP_2 directly rebut each other.
  - Arguments GC_{1a} and GC_{1b} directly rebut each other. Therefore, GC_{1a} also indirectly rebuts arguments MP_2 and DMP_2, namely on GC_{1b}. Likewise, GC_{1b} indirectly rebuts arguments MP_1 and DMP_1, namely on GC_{1a}.

- As for the argument ordering, in my interpretation the court found for Olga Monge on the grounds that the good consequences of adopting the New Rule outweigh the good consequences of adopting the Old Rule. On this interpretation it must be assumed that $GC_{1a} \prec GC_{1b}$, so that GC_{1b} strictly defeats GC_{1a}. Then the argument ordering between the other arguments is irrelevant for the outcome.

It is now easy to see that there is just one preferred labelling (actually displayed in Figure 4). To start with, argument GC_{1b} must be labelled *in* since it has no defeaters (since $GC_{1a} \prec GC_{1b}$). Then GC_{1a} must be labelled *out* since it is directly defeated by an argument that is *in*, namely, GC_{1b}. Then MP_1 is *out* since it has an immediate subargument that is *out*, so DMP_1 is *out* for the same reason. But then DMP_2 must be labelled *in* since its only direct defeater is labelled *out* and none of its subarguments is defeated, so all its immediate subarguments are *in*. In sum, the conclusion that Olga Monge could not be fired for no reason (and so that firing her for no reason was breach of contract) is justified.

7   Conclusions

In this paper the *ASPIC*$^+$ framework for argumentation-based inference was used for formally reconstructing two legal debates about law-making proposals: an opinion of a legal scholar on a Dutch legislative proposal and a US common-law judicial decision on whether an existing common law rule should be followed or distinguished. Both debates were formalised as practical reasoning, that is, as reasoning about what to do. Versions of the argument schemes from good and bad consequences of decisions turned out to be useful in formally reconstructing the debates. This paper has thereby hopefully contributed to clarifying the logical structure of debates about law-making proposals.

Another aim of the case studies was to provide new benchmark examples for comparing alternative formal frameworks for modelling argumentation. Accordingly, an obvious topic for future research is to formalise the same examples in such alternative frameworks and to compare the resulting formalisations with the ones given in this paper.

References

Atkinson, K. and Bench-Capon, T. (2007). Practical reasoning as presumptive argumentation using action based alternating transition systems, *Artificial Intelligence* **171**: 855–874.

Bench-Capon, T. and Prakken, H. (2010). A lightweight formal model of two-phase democratic deliberation, *in* R. Winkels (ed.), *Legal Knowledge and Information Systems. JURIX 2010: The Twenty-Third Annual Conference*, IOS Press, Amsterdam etc., pp. 27–36.

Bench-Capon, T., Prakken, H. and Visser, W. (2011). Argument schemes for two-phase democratic deliberation, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, ACM Press, New York, pp. 21–30.

Besnard, P. and Hunter, A. (2008). *Elements of Argumentation*, MIT Press, Cambridge, MA.

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

Caminada, M. (2006). On the issue of reinstatement in argumentation, *Proceedings of the 11th European Conference on Logics in Artificial Intelligence (JELIA 2006)*, number 4160 in *Springer Lecture Notes in AI*, Springer Verlag, Berlin, pp. 111–123.

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

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

Hage, J. (1997). *Reasoning With Rules. An Essay on Legal Reasoning and Its Underlying Logic*, Law and Philosophy Library, Kluwer Academic Publishers, Dordrecht/Boston/London.

Hage, J. (2004). Comparing alternatives in the law. Legal applications of qualitative comparative reasoning, *Artificial Intelligence and Law* **12**: 181–225.

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

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

Modgil, S. and Prakken, H. (2010). Reasoning about preferences in structured extended argumentation frameworks, *in* P. Baroni, F. Cerutti, M. Giacomin and G. Simari (eds), *Computational Models of Argument. Proceedings of COMMA 2010*, IOS Press, Amsterdam etc, pp. 347–358.

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

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

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

Prakken, H. (2012a). Formalising a legal opinion on a legislative proposal in the ASPIC+ framework, *in* B. Schafer (ed.), *Legal Knowledge and Information Systems. JURIX 2012: The Twenty-fifth Annual Conference*, IOS Press, Amsterdam etc., pp. 119–128.

Prakken, H. (2012b). Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and Dungean semantics, *Artificial Intelligence and Law* **20**: 57–82.

Prakken, H. (2013). Relating ways to instantiate abstract argumentation frameworks, *in* K. Atkinson, H. Prakken and A. Wyner (eds), *From Knowledge Representation to Argumentation in AI, Law and Policy Making. A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday*, College Publications, London, pp. 167–189.

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

Walton, D. (1996). *Argumentation Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, Mahwah, NJ.

Walton, D., Reed, C. and Macagno, F. (2008). *Argumentation Schemes*, Cambridge University Press, Cambridge.
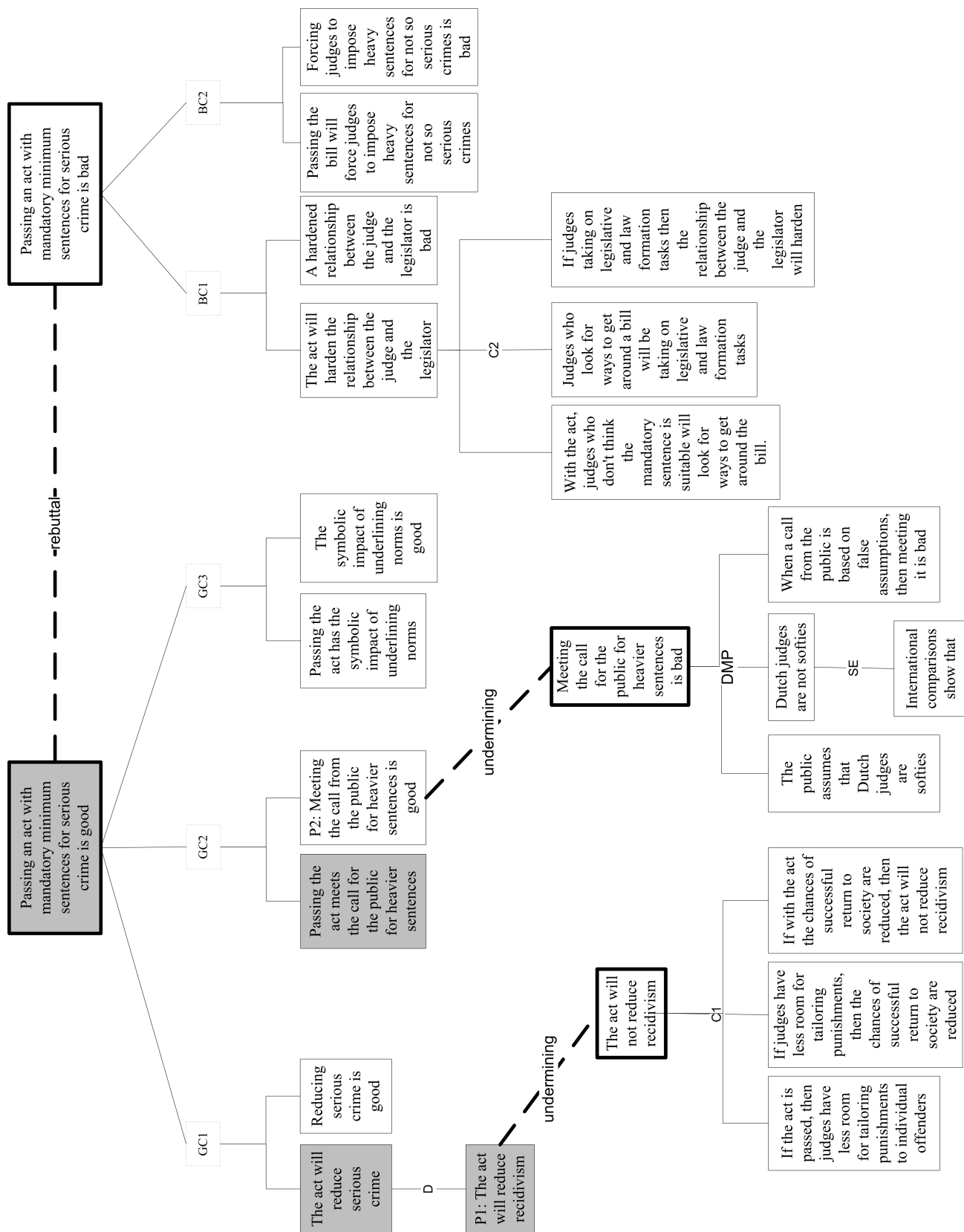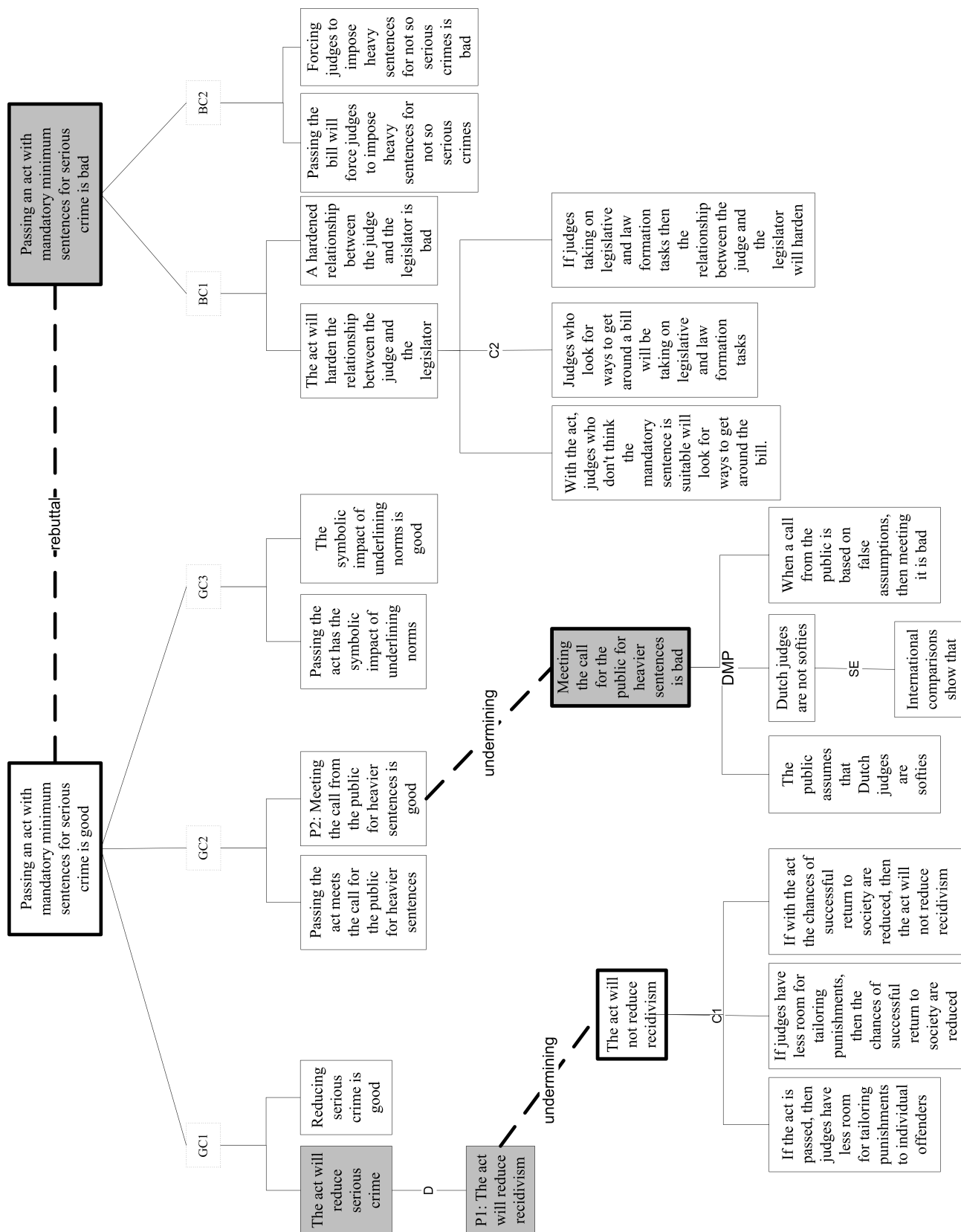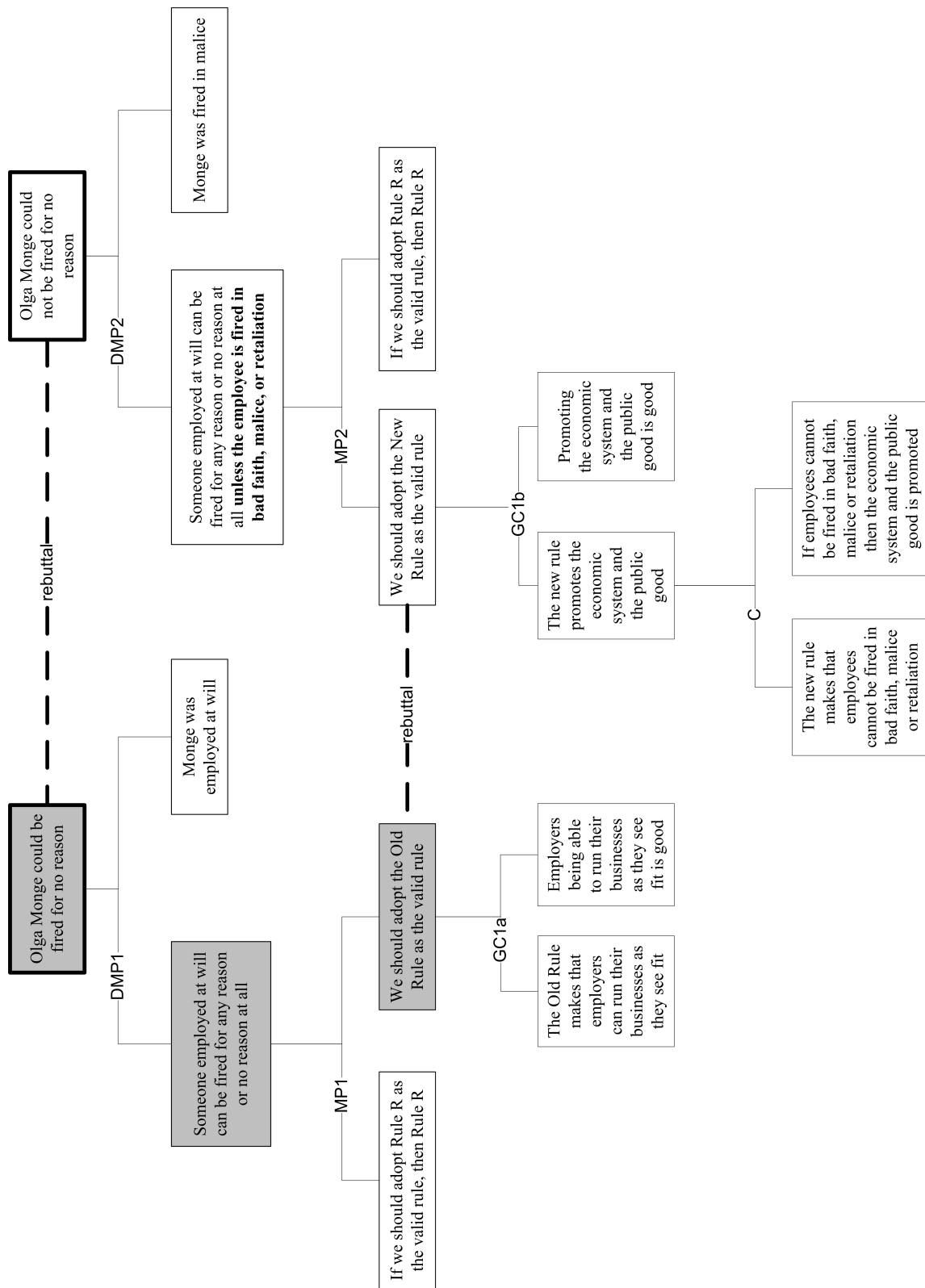
Figure 2: The reconstruction

Figure 3: The second preferred labelling

Figure 4: The second preferred labelling