Formalising a legal opinion on a legislative proposal in the $ASPIC^+$ framework

Henry Prakken

Department of Information and Computing Sciences, University of Utrecht and Faculty of Law, University of Groningen, The Netherlands

Abstract.
This paper presents a case study in which an opinion of a legal scholar on a legislative proposal is formally reconstructed in the $ASPIC^+$ framework for argumentation-based inference. The reconstruction uses a version of the argument scheme for good and bad consequences that does not refer to *single* but to *sets* of consequences, in order to model aggregation of reasons for and against proposals. The case study is intended to contribute to a comparison between various formal frameworks for argumentation by providing a new benchmark example. It also 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.

1. Introduction

In both general AI and AI & law several formal frameworks for argumentation-based inference have been proposed, such as assumption-based argumentation [5], classical argumentation [4], Carneades [7] and $ASPIC^+$ [8]. This raises the question which framework is best suited for formalising natural, in particular legal arguments. The present paper contributes to this discussion with a case study in which an opinion of a legal scholar on a legislative proposal is reconstructed in $ASPIC^+$. While this cannot decide which framework is the best, it helps 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, a main aim of the present case study is to illustrate the usefulness of these features. Defeasible rules will be used to formulate the relevant argument schemes, while argument orderings will be used to apply [6]'s abstract argumentation frameworks for evaluating the justification status of arguments and their conclusions.

The main argument schemes used in the case study are those of good and bad consequences of actions as proposed in [2, 3]. 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 [2, 3] 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. The *ASPIC*$^+$ framework has been applied earlier in a realistic case study in [9]; in that paper the main arguments were not about legislative proposals but about interpreting and applying legal concepts.

This paper is organised as follows. In Section 2 abstract argumentation frameworks and *ASPIC*$^+$ are reviewed. Then in Section 3 the Dutch legal opinion is presented, which is reconstructed in *ASPIC*$^+$ in Section 4. The paper concludes in Section 5.

2. The formalisms

In this section we review abstract argumentation frameworks and the *ASPIC*$^+$ framework. An *abstract argumentation framework* (AF) is a pair $\langle Args, Def \rangle$, where $Args$ is a set of *arguments* and $Def \subseteq Args \times Args$ is a binary relation of *defeat*. A semantics for AFs returns sets of arguments called *extensions*, which are internally coherent and defend themselves against attack. One way to characterise the various semantics is with *labellings*, which assign to zero or more members of $Args$ either the label *in* or *out* (but not both) satisfying the following constraints:

1. an argument is *in* iff all arguments defeating it are *out*.
2. an argument is *out* iff it is defeated by an argument that is *in*.

Stable semantics labels all arguments, while *grounded semantics* minimises and *preferred semantics* maximises the set of arguments that are labelled *in*. In this paper preferred semantics is used, since it allows for alternative labellings and thus for alternative coherent positions. Relative to a semantics, an argument is *justified* on the basis of an AF 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 neither justified nor overruled.

The *ASPIC*$^+$ framework [8] gives structure to Dung's arguments and defeat relation. 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 applying any of [6] semantics for abstract argumentation frameworks to the resulting set of arguments with its defeat relation.

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[1], 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 wff in $\mathcal{L}$ which says that rule $r \in \mathcal{R}$ is applicable. *ASPIC*$^+$ as a framework does not make any assumptions on how the elements of an argumentation system are defined. In *ASPIC*$^+$ argumentation systems are applied to knowledge bases to generate arguments and

[1] In most papers on *ASPIC*$^+$ negation can be non-symmetric, an idea taken from [5]. In this paper we present the special case with symmetric negation.

counterarguments. Combining these with an argument ordering results in argumentation theories, which generate Dung-style AFs.

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 ($\neg$).
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set 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 wff in $\mathcal{L}$), and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $n : \mathcal{R}_d \longrightarrow \mathcal{L}$ is a naming convention for defeasible rules.

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. In what follows, for a given argument the function `Prem` returns all its premises, `Conc` returns its conclusion and `Sub` returns all its sub-arguments.

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

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$;
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 $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$, $\texttt{Conc}(A) = \psi$, $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.

Definition 4 [**Argumentation theories**] An *argumentation theory* is a triple $AT = (AS, KB, \preceq)$ where AS is an argumentation system, KB is a knowledge base in AS and $\preceq$ is an ordering on the set of all arguments that can be constructed from KB in AS (below denoted by $\mathcal{A}_{AT}$). $A \preceq B$ means that B is at least as preferred as A. As usual, $A \prec B$ is defined as $A \preceq B$ and $B \not\preceq A$ and $A \approx B$ as $A \preceq B$ and $B \preceq A$.

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 5 [**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 $B' = \varphi$, $\varphi \notin \mathcal{K}_n$.

Undercutting attacks succeed as *defeats* independently of preferences over arguments, since they express exceptions to defeasible inference rules. Rebutting and undermining attacks succeed only if the attacked argument is not stronger than the attacking argument.

Definition 6 [**Defeat**] A *defeats* B iff:A undercuts B, or; A rebuts/undermines B on B' and $A \nprec B'$. A strictly *defeats* B iff A *defeats* B and B does not *defeat* A

Abstract argumentation frameworks are then generated as follows:

Definition 7 [**Argumentation frameworks**] An *abstract argumentation framework (AF) corresponding to an* $AT = \langle AS, KB, \preceq \rangle$ is a pair $\langle Args, Def \rangle$ such that:

- $Args$ is the set $\mathcal{A}_{AT}$ as defined by Definitions 3 and 4,
- Def is the relation on $Args$ given by Definition 6.

Now 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.

3. 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, published on 29 February 2012 at www.rug.nl/rechten. 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.'

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.'

4. 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 [3] and some other schemes. Contrary to the usual formulations of schemes from consequences (e.g. [10, 1]), they do not refer to single but to *sets of* good or bad consequences.[2]

Argument scheme from good consequences

$$\begin{array}{l} \text{Action } A \text{ results in } C_1 \\ \ldots \\ \text{Action } A \text{ results in } C_n \\ C_1 \text{ is good} \\ \ldots \\ C_n \text{ is good} \\ \hline \text{Action } A \text{ is good.} \end{array}$$

Argument scheme from bad consequences

$$\begin{array}{l} \text{Action } A \text{ results in } C_1 \\ \ldots \\ \text{Action } A \text{ results in } C_m \\ C_1 \text{ is bad} \\ \ldots \\ C_m \text{ is bad} \\ \hline \text{Action } A \text{ is bad.} \end{array}$$

These schemes have three critical questions:

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

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

In *ASPIC*$^+$ these questions are pointers to counterarguments. Question 1 points to underminers, question 2 to rebuttals and question 3 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.

My reconstruction of Kwakman's opinion is visualised in Figure 1. 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 the figure 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 causes Q, Q is by definition a case of R strictly implies P causes R.
- C1 and C2 stand for two applications of causal chaining: P_1 causes P_2, P_2 causes ... causes P_n strictly/defeasibly implies P_1 causes P_n (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 defeasibly implies Q.
- SE is shorthand for a 'scientific evidence' scheme: scientific evidence shows that P defeasibly implies P.

The links in Figure 1 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 in dotted boxes for any of these reasons, and then linking these dotted boxes to the conclusion that A is good. 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 $GC123$), three with any combination of two reasons (denoted below by $GC12, GC13, GC23$) and three applying any individual reason (denoted below by $G1, G2$ and $G3$).

The argumentation theory corresponding to Figure 1 can be summarised as follows:

- $\mathcal{L}$ is a first-order 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.
- 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 two argument trees (where their conclusions are regarded as their roots). $\mathcal{K}$ thus consists of 18 ordinary premises.

The argumentation framework induced by this argumentation theory 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.
- The attack relations are more in number than the three shown in the figure:
  * 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 that $P_1 \prec C_1$ while $DMP \approx P_2$. The resulting defeat graph is shown in Figures 2 and 3. Here mutual defeat relations are for ease of readability displayed as dotted arrows, while premise arguments are omitted if they are not attacked, and BC_1 and BC_2 are omitted since the focus is on whether any argument in favour of the legislative proposal can be adopted.

The graphs show two preferred labellings, where white means that the argument is *in* and grey that it is *out*. Note that in both labellings all GC-arguments using premise P_1 are *out*, since they are (indirectly) strictly defeated by C_1, which (directly) strictly defeats the premise argument P_1. The two labellings then differ on how they (arbitrarily) resolve the two mutual defeat relations, namely between GC_{23} and BC_{12} and between DMP and P_2. The labelling in Figure 2 results from accepting GC_{23} and P_2 at the expense of BC_{12} and DMP. Note that GC_{23} can only be accepted if P_2 is also accepted, otherwise DMP makes GC_{23} *out*. The labelling in Figure 3 results from accepting BC_{12} and DMP at the expense of GC_{23} and P_2. Note that BC_2 can be accepted without accepting DMP, so this defeat graph has a third labelling, which is equal to the one in Figure 3 except that DMP is *out* and P_2 is *in*.

5. Conclusions

In this paper I have illustrated the potential of the *ASPIC*$^+$ framework for argumentation-based inference as a tool for reconstructing natural legal argument about legislative proposals. The case study illustrated that argument schemes can be conveniently modelled as defeasible inference rules and that *ASPIC*$^+$'s notion of an argument ordering in combination with [6]'s semantics for abstract argumentation frameworks provides a suitable way to evaluate debates. The case study also suggests that modelling legal policy arguments does not always require a distinction between goals and values, as e.g. modelled in [1].

One aim of the case study was to provide a new benchmark example for comparing alternative formal frameworks. Accordingly, an obvious topic for future research is to formalise the same example in such alternative frameworks and to compare the resulting formalisations with the one given in this paper.

References

[1] K.D. Atkinson and T.J.M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171:855–874, 2007.

[2] T.J.M. Bench-Capon and H. Prakken. A lightweight formal model of two-phase democratic deliberation. In R.G.F. Winkels, editor, *Legal Knowledge and Information Systems. JURIX 2010: The Twenty-Third Annual Conference*, pages 27–36. IOS Press, Amsterdam etc., 2010.

[3] T.J.M. Bench-Capon, H. Prakken, and W. Visser. Argument schemes for two-phase democratic deliberation. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, pages 21–30, New York, 2011. ACM Press.

[4] Ph. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, Cambridge, MA, 2008.

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

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

[7] T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.

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

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

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

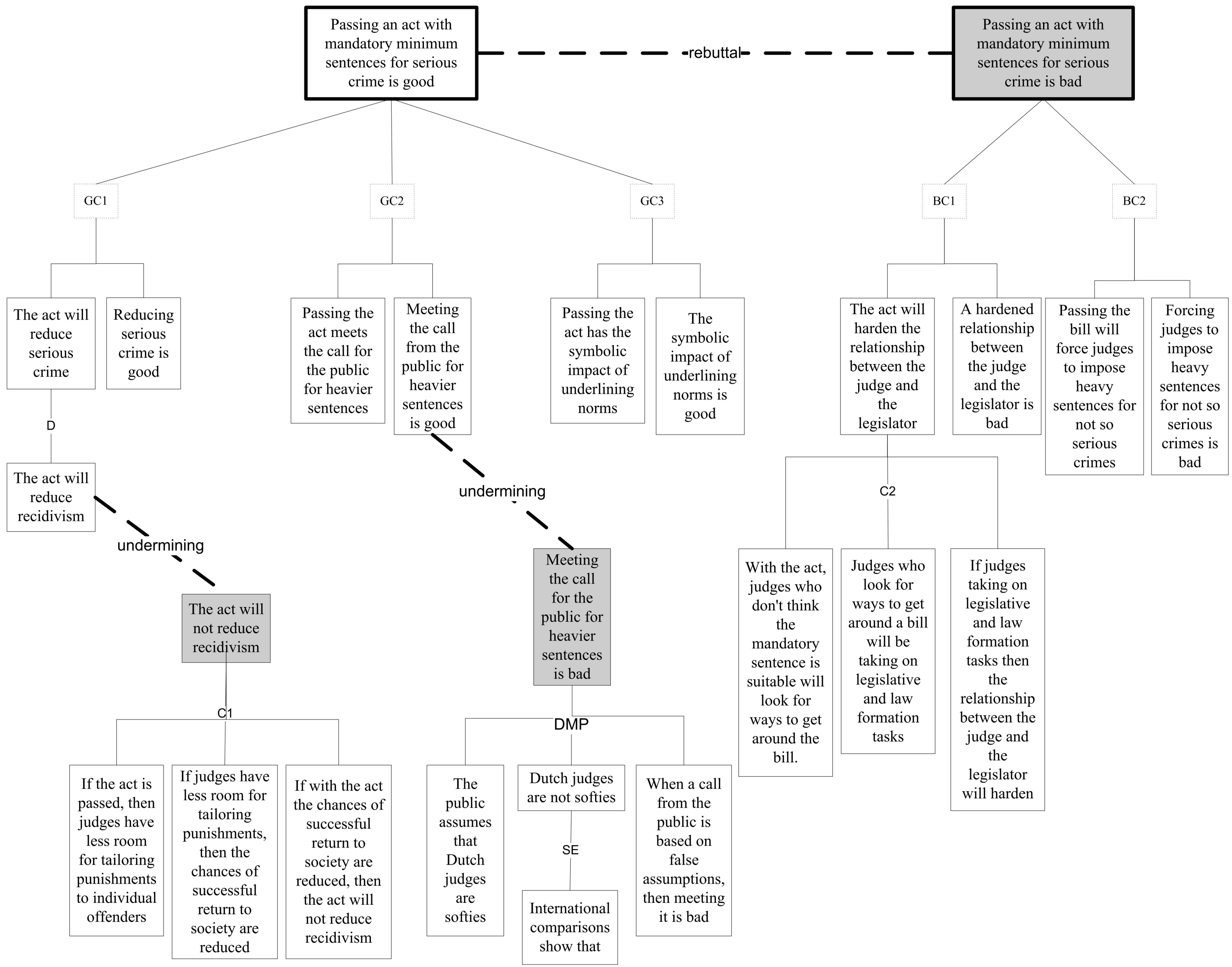

Figure 1.: The reconstruction

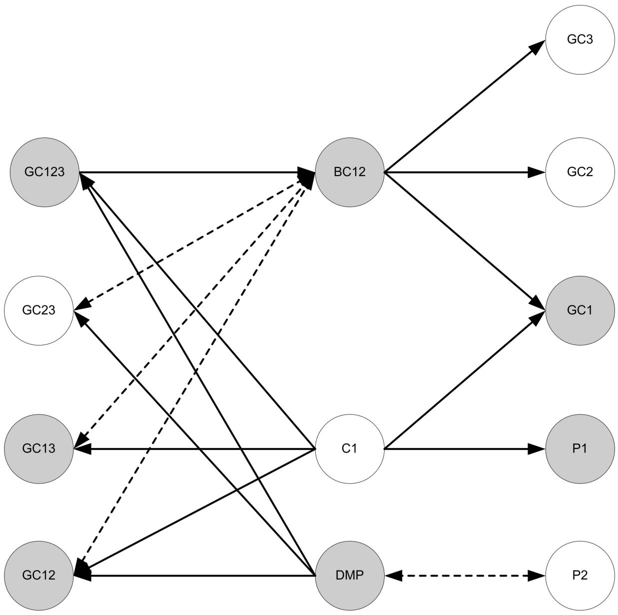

Figure 2.: Defeat graph and preferred labelling

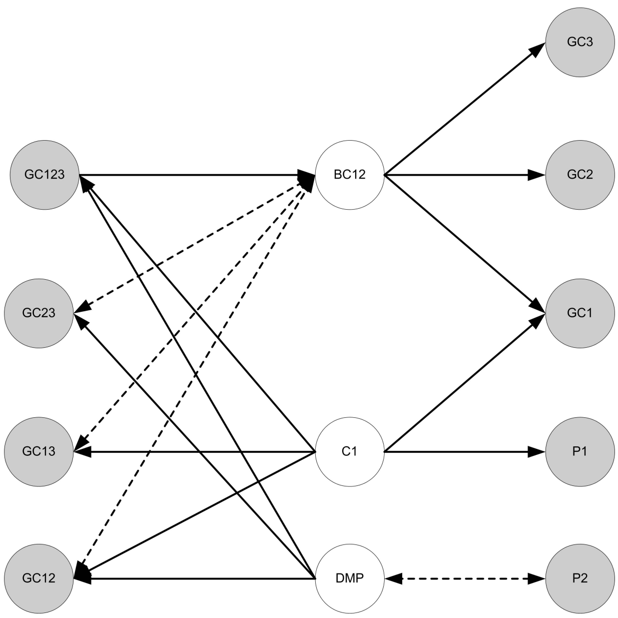

Figure 3.: Defeat graph and alternative preferred labelling