Argument schemes for two-phase democratic deliberation

Trevor Bench-Capon
Department of Computer Science, University of Liverpool
UK

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

Wietske Visser
Man-Machine Interaction Group, Delft University of Technology
The Netherlands

ABSTRACT

A formal two-phase model of democratic policy deliberation is presented, in which in the first phase sufficient and necessary criteria for proposals to be accepted are determined (the 'acceptable' criteria) and in the second phase proposals are made and evaluated in light of the acceptable criteria resulting from the first phase. Such a separation gives the discussion a clear structure and prevents time and resources from being wasted on evaluating arguments for proposals based on unacceptable criteria. Argument schemes for both phases are defined and formalised in a logical framework for structured argumentation. The *process* of deliberation is abstracted from and it is assumed that both deliberation phases result in a set of arguments and attack and defeat relations between them. The acceptability status of criteria and proposals within the resulting argumentation framework is then evaluated using preferred semantics. For cases where preferences are required to choose between proposals, inference rules for deriving preferences between sets from an ordering of their elements are given.

1. INTRODUCTION

Discussions on policy proposals often contain two separate phases: first criteria that proposals should satisfy are determined and then specific proposals are put forward and evaluated against the criteria previously established. Possible benefits of such a separation are that in this way the discussion has a clear structure, that the choice of criteria is not influenced by the proposals put forward, and that no time and resources are wasted on evaluating proposals using unacceptable criteria. For contexts where these benefits are desirable, we present a formal two-phase model of democratic policy deliberation. In less organised contexts, where criteria and proposals are advanced in a less systematic fashion, our model can still provide a useful tool to analyse and evaluate the discussion. In the first phase sufficient and necessary criteria for proposals to be accepted as having desirable features are determined (the 'acceptable' criteria) and in the second phase proposals are made and evaluated in light of their merits as determined by the acceptable criteria resulting from the first phase. The idea is that the criteria are objectively measurable properties of proposals, meant to give substance to the often subjective and abstract desires. We abstract from the *process* of deliberation and so do not consider how the arguments are put forward, but simply assume that both deliberation phases result in a set of arguments and attack and defeat relations between them. We then apply preferred semantics [7] to evaluate the acceptability status of criteria and proposals within the resulting argumentation framework. Preferred semantics is used since this arguably suits the inherently credulous nature of argumentation over action, where often the *best* action is a matter of subjective preference rather a matter of objective fact. Thus, if action proposals are conflicting and the conflict cannot be resolved through logic alone, in the end a choice has to be made. The second phase is constrained by the first phase in that the evaluation of proposals can only make use of the criteria that were accepted in the first phase.

The core of our model is two sets of argument schemes. The schemes for the first phase allow citizens to argue for and against necessary and sufficient criteria. The schemes for the second phase allow arguments for and against proposals to be made and evaluated in terms of how well they satisfy the acceptable criteria resulting from the first phase. The language of the argument schemes is a light-weight formal one, to fit with the intended e-democracy applications.

The dialogue setting we assume is subjective but social. Central to our argument schemes is that criteria can be justified by saying that they are measurable attributes indicating desirable properties of proposals. It is necessary, however, to distinguish what is desirable for particular individuals from what is desirable for the group as a whole. The distinction goes back to Rousseau [14]:

> There is often a great deal of difference between the will of all and the general will. The latter looks only to the common interest; the former considers private interest and is only a sum of private wills.

Ideally in a democracy people will indeed vote for what they consider to be the common good rather than from selfish motives. Of course, in practice this may not always be the case, as Rousseau [14] recognised:

> Finally, when the State, on the eve of ruin, maintains only a vain, illusory and formal existence, when in every heart the social bond is broken, and the meanest interest brazenly lays hold of the sacred name of "public good," the general will becomes mute: all men, guided by secret motives, no more give their views as citizens than if the State had never been; and iniquitous decrees directed solely to private interest get passed under the name of laws.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICAIL '11 June 6-11, Pittsburgh, USA
Copyright 2011 ACM 978-1-4503-0755-0/11/06 ...$10.00.

Often the truth lies between the two, with people torn between endorsing what is good for society as a whole and their own self interest. Often too there are differences between what people say in public and what they vote for in a secret ballot: income tax rises to fund services for the poorest receive far more support in opinion polls than at the ballot box. None the less we will in this paper assume that the users of our model are well intentioned and will deliberate about what is good for society as a whole and not about what is good for themselves, leaving individual goods for future work. We are aware that this is an idealised position, but we believe it is good to start with an idealised model and then see how it can be enriched. In our argument schemes therefore *it is desirable* is to be understood as *it is desirable for society as a whole*, and cannot be justified simply by some person or group actually desiring it.

Even if one regards this as too high minded for a tool to support public debate, it is certainly the setting in which policy questions are deliberated by the public officials who support political decision makers. Such officials are intended to be politically neutral, and to formulate policies which will satisfy the desires, and the priorities, of their political masters. Intra-government debates can thus be seen as following these principles.

This paper significantly revises and extends work reported in [4]. In brief some improvements and simplifications have been made and an entirely new set of inference rules for deriving premises has been added. A detailed comparison with [4] will be given in the concluding section.

This paper is organised as follows. In section 2 we briefly describe the formal setting: argumentation frameworks, an extension to allow the expression of preferences, and a means of providing structure for arguments. Section 3 describes the phase in which criteria are proposed and established. Section 4 describes the second phase in which proposals are evaluated against these criteria. Section 5 illustrates the application of the model with two examples and section 6 offers some discussion and conclusions.

2. THE FORMAL SETTING

We first briefly summarise the three formal frameworks we will use in this paper.

Dung's *abstract argument frameworks* [7] are a pair $AF = \langle \mathcal{A}, defeat \rangle$, where $\mathcal{A}$ is a set of arguments and *defeat* a binary relation on $\mathcal{A}$. A subset $\mathcal{B}$ of $\mathcal{A}$ is called *conflict-free* if no argument in $\mathcal{B}$ defeats an argument in $\mathcal{B}$ and it is called *admissible* if it is conflict-free and defends itself against any attack, i.e., if argument A_1 is in $\mathcal{B}$ and argument A_2 defeats A_1, then some argument in $\mathcal{B}$ defeats A_2. A *preferred extension* is then a maximally (wrt set inclusion) admissible set. Dung defines several other types of extensions but for our model they are not needed.

Modgil's [10] *extended argumentation frameworks* refine those of Dung in two ways. First, instead of a defeat relation Modgil assumes a more basic *attack* relation, standing just for notions of syntactic conflict. Then Modgil allows *attacks on attacks* in addition to attacks on arguments. Intuitively, if argument C claims that argument B is preferred to argument A, and A attacks B, then C undermines the success of A's attack on B (i.e., A does not *defeat* B) by pref-attacking A's attack on B. Thus an *extended argumentation framework* is a triple $EAF = \langle \mathcal{A}, attack, pref\text{-}attack, \rangle$, where $attack \subseteq \mathcal{A} \times \mathcal{A}$ and $pref\text{-}attack \subseteq \mathcal{A} \times attack$. Then defeat is made relative to a set of arguments: for any subset S of $\mathcal{A}$ and arguments A and B: B *S-defeats* A iff B attacks A and there is no argument C in S that pref-attacks this attack. Dung's theory of AFs is then reformulated with *defeat* replaced by *S-defeat*. Since arguments attacking attacks can themselves be attacked, as can these attacks, and so on, Modgil's extended argumentation frameworks can fully model argumentation about whether an argument defeats another.

Another refinement of Dung's abstract approach is the ASPIC framework for structured argumentation [13]. This framework assumes an unspecified logical language and defines arguments as inference trees formed by applying inference rules (which may be either strict or defeasible) to a knowledge base: the nature of the inference rules is also unspecified. The notion of an argument as an inference tree leads to three ways of attacking an argument: attacking an inference (*undercut*), attacking a conclusion (*rebuttal*) and attacking a premise (*undermining*), where rebutting and undercutting attacks can only be targeted at applications of *defeasible* inference rules. To resolve underminings and rebuttals, a preference relation $\prec$ on arguments (to be specified as input) is used, which leads to three corresponding kinds of defeat: undercutting, rebutting and undermining defeat. Basically, A successfully rebuts (undermines) B if A rebuts (undermines) B and $A \not\prec B$. Then A defeats B if A undercuts B or successfully rebuts or undermines B. Note that this means that undercutting attacks always succeed, irrespective of preferences.

Finally, in [11] EAFS are instantiated with the ASPIC framework, where ASPIC's input ordering $\prec$ on arguments is replaced by pref-attacks on attacks. This results in a logical framework for structured argumentation with attacks on attacks (below called E-ASPIC), so that now arguments can be built that make explicit why an attack is attacked. Thus while an argument attacking an attack may simply be the expression of preference, it could also be a justification of that preference. Importantly admitting such arguments to the framework allows for them to be attacked by arguments expressing or justifying the contrary preference. In the remainder of this paper we will use E-ASPIC as the logical framework.

3. THE CRITERIA DELIBERATION PHASE

In the criteria deliberation phase the participants propose and attack possible criteria for assessing the proposals to be made in the second phase. The aim is to identify objective measurable attributes which will indicate the, typically less tangible, desiderata which the proposals attempt to realise. In this section we give a set of argument schemes (defeasible inference rules in the E-ASPIC framework) that can be used for these purposes. The distinction between sufficient and necessary criteria (relative to something that is desirable) is important, as these will be used differently in the second phase. The outcome of this phase is a set of acceptable criteria; this set will be used in the second phase to make and assess proposals.

The first phase assumes a given set of social desires, some of which may have been declared essential in that all proposals must satisfy them. This establishes the context for the deliberation: the obligatory goals which all proposals must achieve, and the optional goals which enable us to choose between proposals which satisfy the basis requirements. In an extended model the given desires may themselves be the outcome of an argumentation process, but in the present model we abstract from that, and do not discuss how the desiderata are established. (Note, however, that our embedding of the argument schemes in the E-ASPIC framework allows a straightforward extension of the model in this respect.) Given the desires, the participants propose criteria for proposals in terms of these desires. Recall that these criteria are meant to be objectively measurable properties of proposals. Logically they are of two kinds (see also [17]). Sufficient criteria are such that if they are satisfied, at least one of the desires is realised. Note that this makes a sufficient criterion relative to a given desire: what is sufficient to realise one desire may not be sufficient to realise another desire or may even prevent the realisation of another desire. Necessary criteria are also relative to given desires: they are such that if they are not satisfied,

at least one desire is violated.

The distinction between necessary and sufficient criteria is essentially the familiar logical distinction between 'if' and 'only if'. For example, 'If Gross National Product rises then prosperity' says that a rise in GNP is a sufficient criterion for prosperity, so that a rise in GNP ensures that a country is prosperous, however that additional wealth is distributed. In contrast, 'If prosperity then Gross National Product increases' says that a rise in GNP is a necessary criterion for prosperity, so *only if* GNP rises can prosperity be achieved. This is the same as 'no country without a rising GNP is prosperous'. ('If prosperity then GNP rises' is equivalent to 'If no rise in GNP then not prosperous', so any country that without a rising GNP is not prosperous.) This seems reasonable: if there is to be prosperity there has to be some additional wealth to share out. But note also that if a rise in GNP is only a necessary condition for prosperity, then there can still be countries with a rise in GNP which are not prosperous. Inequality may also rise, so that all the additional, and some of existing, wealth goes to a small number of people, leaving the majority worse off than before.

3.1 Argument schemes for determining criteria

Sufficient criteria can be proposed with a variant of [18]'s scheme from good consequences ("if P is brought about, then good consequences will occur, therefore P should be brought about"). Moreover, because of the equivalence of 'if desire then criterion' and 'if not criterion then not desire', necessary criteria can be proposed with a variant of Walton's scheme from bad consequences ("if P is brought about, then bad consequences will occur, therefore P should not be brought about"). The main differences from Walton's schemes are that our conclusions are not proposals for actions but suggested criteria for ascribing consequences to action proposals so that our conclusions are made relative to the respect in which the consequences are good which motivates the conclusion.

SCS: C should be a sufficient criterion for D since C satisfies D and D is desirable

NCS: C should be a necessary criterion for D since D requires C and D is desirable

Note that by the equivalence of 'D requires C' and 'Not-C results in Not-D' the *NCS* scheme can also be used to deal with negative side-effects of an action. Note also that this says that the criteria is a sufficient, respectively necessary, *sign of D*, and should not be confused with it being a sufficient, respectively necessary, condition for adopting a proposal. Where a desire has been designated *essential*, however, a criterion necessary for that desire becomes a condition necessary for a proposal to be adopted. This will be reflected in our argument schemes for the second phase. A sufficient criterion for an essential desire, however, is *not* a sufficient condition for a proposal to be adopted, since there may other essential desires to consider, and there may be several proposals satisfying that criterion.

The remaining schemes generate rebutting attacks on uses of these two schemes. At first sight it would seem that if two necessary criteria are incompatible, the arguments supporting them should attack one another. For example, if it is suggested that sound economic management requires a balanced budget and it is alternatively suggested that sound economic management requires a deficit budget, it would seem at first sight as if the arguments suggesting these contentions should attack each other. However, upon closer inspection it turns out that this only holds if the two criteria are suggested *for the same desire*. In that case the arguments should surely attack each other since there is a difference of opinion on what is required for economic management to be sound in respect of that desire. In fact, of course, there are many aspects to sound economic management. This means that the necessary criterion that there be a budget deficit can be motivated by the desire to reduce unemployment, while the desire for a balanced budget is to keep inflation low. Then there should be no logical conflict since the necessary criteria are relative to different desires, namely, to have low unemployment and to have low inflation, both of which contribute in their own way to sound economic management. Of course, there will be no proposal that satisfies both criteria, but this situation will be dealt with in the second phase, where proposals are constructed and compared, and a choice made as to which aspect should be given priority in the particular situation and with respect to the available actions. The proposals designed to reduce unemployment through a budget deficit by spending on infrastructure and designed to reduce inflation through balancing the budget by cutting rail subsidies will then be compared in terms of a preference ordering on the motivating desires (having low unemployment versus having low inflation), and the other desires which those proposals realise or fail to realise. To enable such a comparison in the second phase, the output of the first phase should regard both 'balanced budget' and 'budget deficit' as acceptable criteria for their respective desires. The same analysis holds for incompatible necessary and sufficient criteria: they only conflict if they are for the same desire: in that case an opinion on how a desire can be satisfied conflicts with an opinion on what is required for satisfying the desire. Note that incompatible sufficient criteria do not conflict even if they are for the same desire, as they are non-exclusive alternatives rather than rivals. For example, both a balanced budget and a budget surplus may be sufficient for low inflation.

In fact, since conflicts between necessary and/or sufficient criteria boil down to conflicts about how a given desire can or must be satisfied, the conflict schemes will not refer to the criteria at all but to the 'requires' and 'satisfies' statements that motivate them. Together these observations result in the following three conflict schemes, which also deal with the possibility that the conflict is between more than two statements (here the compatibility relation is assumed to be symmetric).

CN2N: D does not require C since D requires C_1 and ... and D requires C_n and C is not compatible with all of $C_1, \ldots, C_n$.

CN2S: C does not satisfy D since D requires C_1 and ... and D requires C_n and C is not compatible with all of $C_1, \ldots, C_n$.

CS2N: D does not require C since C_1 satisfies D and D requires C_2 and ... and D requires C_n and C is not compatible with all of $C_1, \ldots, C_n$.

Only one sufficient criterion is needed in *CS2N*, since if more than one sufficient criterion were required to generate an incompatibility with C, we could still have acceptable proposals satisfying all the proposed necessary and one or other of the sufficient criteria, and different proposals could satisfy the different sufficient criteria. Thus this would not exclude any criterion.

Let us see how these schemes can be used to attack applications of the *SCS* or *NCS* scheme. For such attacks two questions arise: are these attacks symmetric and can preferences be used to determine whether an attack is successful, i.e., whether the attack always results in defeat? Note that these questions are independent of each other, since in E-ASPIC, and other argumentation systems using preferences, an asymmetric attack of A on B is unsuccessful if B is preferred over A.

Consider, for example the following two arguments (with the obvious abbreviations):

$A_1 = NC(C_1)$ since D requires C_1 and D is desirable
$B_1 = NC(C_2)$ since D requires C_2 and D is desirable

Then if it is given that C_1 and C_2 are incompatible with each other, the following attacks can be constructed.

$A_2 = D$ does not require C_2 since D requires C_1 and C_2 is not compatible with C_1
$B_2 = D$ does not require C_1 since D requires C_2 and C_1 is not compatible with C_2

Suppose furthermore that the 'requires' premise of A_1 is the conclusion of an unspecified defeasible subargument A'_1 and that likewise the 'requires' premise of B is the conclusion of an unspecified defeasible subargument B'_1. Then we have that A_2 rebuts B_1 at B'_1 and B_2 rebuts A_1 at A'_1. Note that A'_1 is also a subargument of A_2 and B'_1 is also a subargument of B_2, so A_2 also rebuts B_2 at B'_1 and B_2 also rebuts A_2 at A'_1. Summarising, we have the following attacks:

A_2 rebuts B_2 on B'_1
B_2 rebuts A_2 on A'_1
A_2 rebuts B_1 on B'_1
B_2 rebuts A_1 on A'_1

Intuitively these attacks express just one conflict, which should be resolved by comparing the arguments A'_1 and B'_1, since these have conflicting conclusions on what is required by D.

In the E-ASPIC framework this result can be obtained as follows. Firstly, all conflict schemes should be formalised as defeasible inference rules (this seems reasonable if the 'satisfies' and 'requires' statements are meant to express what is typically the case, allowing for exceptions). Secondly, the strength of applications of the conflict schemes should be defined in terms of the strength of the subarguments for their 'requires' and 'satisfies' premises. Note that this is natural since in practice the incompatibility premises of the conflicting arguments will always be two sides of the same symmetric incompatibility relation, so these premises will always have equal priority. Then A_2 is exactly as strong as A'_1 and B_2 is exactly as strong as B'_1, so all conflicts are effectively resolved by comparing A'_1 and B'_1 as desired.

If the incompatibility involves more than two criteria, then the analysis is more complicated but yields a similar outcome. Consider, for instance, a conflict that is essentially between three 'requires' statements, provided by defeasible arguments A, B and C. Then three pairwise comparisons must be made:

A versus $\{B,C\}$
B versus $\{A,C\}$
C versus $\{A,B\}$

With any reasonably defined argument ordering the argument using the weakest of these three arguments will then be defeated. We leave it to the reader to verify the details of this analysis.

While this formalisation in E-ASPIC yields the intended outcomes, it formally multiplies an attack that intuitively is a single one. It would be interesting to investigate how the E-ASPIC framework could be extended with arguments about whether arguments rebut each other but this has to be left to a future occasion and is more of a technical matter for ASPIC. For our purposes here, it is enough that a single preference resolves the conflict.

Finally, recall that applications of all schemes can be attacked on their premises. For example, a dispute as to what is desirable would result in arguments attacking and defending the premise of *SCS* or *NCS* that D is desirable. Moreover, the last premise of the three conflict schemes allows a debate to be about whether a set of criteria is compatible. Sometimes the conflict may be logical: to *Save Money* and to *Give Fiscal Stimulation* can both be criteria for prudent economic management, but fiscal stimulation just is spending money. Alternatively the arguments for conflict may offer contextual reasons why criteria cannot be jointly satisfied, such as 'we cannot appoint a committee member who is both a woman and disabled even though both are needed to promote diversity, since there are no disabled female applicants', reflecting particular, rather than general, circumstances. Debates about what criteria should be used to establish that a desire is satisfied, and about what is desirable are an important feature of political debate, and our model allows for any set of argument schemes for attacking or supporting premises of the above schemes. This, however, can use any techniques of general argumentation and so, in this paper, where we focus on argumentation *specific* to these debates, we will not go into this, except for an extended informal example in Section 5.1.

3.2 The outcome of the criteria deliberation phase

With the schemes proposed above (plus possibly other schemes for supporting or attacking their premises) an extended argumentation framework in the sense of [11] is built. At the end of the first phase, the preferred extensions are determined. When there are attack relations between arguments that cannot be resolved in favour of one of the arguments, multiple extensions result. Given the analysis thus far, this can happen in the following cases:

- disagreement about what should be desirable
- disagreement about which desires are essential
- disagreement about how desires can or must be satisfied
- disagreement about incompatibility of criteria.
- Recursively, any kind of disagreement with respect to conclusions or premises of subarguments of arguments on the first four issues.

From this it follows that in the first phase preferences on desires are irrelevant, so in this phase no arguments on such preferences will be stated. However, any other kind of preference may be relevant. It is even possible that some conflicts between arguments are decided by a voting procedure; the result of a vote can in E-ASPIC be expressed as a simple preference argument consisting just of a preference statement. We could require our participants to resolve all the various issues, and so choose between the preferred extensions. Note that because we have no conflict depending on preferences over desires, and because desirability is intended to be a objective matter for the participants, it is not unreasonable to expect them to be able to resolve the issues, either by finding more information, or by voting.

If we do not insist on this and so allow multiple preferred extensions the acceptable criteria for the proposal deliberation phase must still be identified (some of which may have been determined to be essential). Here there are two possibilities:

1. Only criteria that are conclusions in each extension are acceptable. This results in a unique set of acceptable criteria.
2. All criteria that are conclusions in at least one extension are acceptable. This also results in a unique set of acceptable criteria, although some of the criteria may be based on conflicting points of view. It means that every defensible criterion can be considered.

3. Each preferred extension results in an alternative acceptable-criteria set and each proposal must be made relative to such a set.

Clearly the first method is simpler. Because it insists on skeptical acceptance for criteria, a desire which is the subject of an unresolved incompatibility may have no criterion at all in the second phase, even though its desirability is unquestioned. If what should count as the desire being satisfied really cannot be agreed, this may be what we want. Other methods ensure that the desire is considered in phase two, but may only postpone the issue of which criteria set should be used until that phase. This is undesirable because then the consequences of the choice in terms of the proposal that will be accepted will be known, which may bias the decision as to the criterion to use.

In this paper we assume that a method giving a single acceptable set is used, that is, either the first or the second method, but our choice here is not significant, since if desired, our analysis below can be easily adapted to the third method in the way of [4].

4. THE PROPOSAL DELIBERATION PHASE

In the proposal deliberation phase action proposals can be made and supported by sets of sufficient criteria that they satisfy, where such sets must at least contain sufficient criteria for all essential desires. Proposal arguments can be attacked in two ways.

- They can be attacked on their premises by arguments claiming that the proposal does not satisfy some given sufficient criterion; whether this attack is symmetric depends on the nature of the attack.
- By alternative proposal arguments. These arguments can either take their premises from the same or from another acceptable criterion set. This kind of attack is symmetric.

Note that in these attacks necessary criteria are not utilised. Necessary criteria, because they represent constraints on, rather than reasons for, action, are instead used in arguments that pref-attack attack relations. Such arguments summarise for each of the two conflicting proposal arguments which sets of sufficient conditions they satisfy and which sets of necessary conditions they violate, and they then express a preference between the proposals based on preferences on the desires motivating these criteria.

Note also that a proposal argument cannot simply be attacked by pointing at a sufficient criterion that it does not satisfy or a necessary criterion that it violates. Such attacks are always part of alternative proposal arguments. Allowing such attacks without an alternative proposal would point to defects in proposals without suggesting a better alternative. Since we are looking for the best *available* solution we can have no reason to suppose that any solution will satisfy all our desires. Therefore failure to satisfy a desire is not a reason to reject a proposal, provided that it does realise all essential desires. The only reason to reject a proposal which satisfies all essential desires is that there is a better proposal.

We abstract from the internal structure of proposals: for example, from whether a proposal concerns atomic or combined actions. We thus leave room for proposals that include other proposals (for example, to both raise taxes and cut social benefits). In much other work on argumentation over action (e.g.[2]) it is assumed that only one action can be performed in a situation but for democratic deliberation this assumption is not realistic, as our example shows. If a proposal that combines two actions over commits since one of the actions satisfies the same sufficient criteria, then (if the debate is conducted properly) this will reflect itself in violation of a necessary criterion (such as 'do not put more financial burdens on citizens than necessary').

4.1 Argument schemes for proposal deliberation

The argument scheme for generating proposals has the following form:

PS: proposal P should be adopted since
P satisfies sufficient criterion c_1 for d_1 and
$\ldots$ and
P satisfies sufficient criterion c_n for d_n and
$d_1 \neq \ldots \neq d_n$ and
$\{d_1, \ldots, d_n\}$ includes all essential desires.

The general scheme for preference arguments is as follows:

PrS: proposal P_2 is preferred over proposal P_1 since
P_1 satisfies sufficient criteria for $D_1^+ = \{d_1, \ldots, d_m\}$
P_2 satisfies sufficient criteria for $D_2^+ = \{d_n, \ldots, d_p\}$
P_1 violates necessary criteria for $D_1^- = \{d_q, \ldots, d_r\}$
P_2 violates necessary criteria for $D_2^- = \{d_s, \ldots, d_t\}$
$(D_1^+, D_1^-) < (D_2^+, D_2^-)$

The conclusion of this scheme is assumed to pref-attack P_1's attack on P_2. Like all other schemes, this scheme can be attacked on any of its premises, for example, on whether a proposal really violates some necessary criterion, or on the ordering on sets. In fact, this is the only way in which preference arguments can be attacked, since the idea is that if all premises hold, then the preference conclusion holds by definition. Thus all disagreement and defeasibility is located in the premises of this scheme.

Note that if there is complete knowledge about whether a proposal satisfies any given criterion and if criterion values are always either satisfied or not, it is not necessary to keep track of the desires that are violated. In that case proposals can simply be ranked by looking at which criteria a proposal satisfies, since if it is not known to realise a given desire, it is known to violate it. However, in general this assumption does not hold: for example if a budget surplus is desired, we may well want to treat a small surplus or even a small deficit, as not worth considering, and so see that desire as neither satisfied not violated. The precise threshold can be itself the subject of argumentation. [1] In general, therefore, we must also keep track of the criteria that are violated. This in turn leads to a double complication. Not only must sets be ordered in terms of an ordering on their elements but also must pairs of such sets be compared with each other.

A key issue then is how the last premise of the *PrS* scheme is verified. A crude way is to leave this entirely to the discussants by allowing them to construct any argument for or against the premise. However, there is a considerable literature on combining preferences and fully free preference arguments may violate rationality constraints proposed in this literature. Therefore, an alternative approach is to only allow the discussants to construct preference argument concerning individual desires and to define inference rules for combining these preferences that formalise a suitable method from the literature. We will explore this approach in the next subsection.

[1] In *David Copperfield* by Charles Dickens there is a very famous saying of Mr Micawber which expressed the difference between happiness and misery as the difference between a budget surplus or deficit of sixpence (four euro cents). A surplus or deficit of three pence would doubtless have left him indifferent.

$$1 \quad \frac{|D^+_{1,i}| = n \quad |D^-_{1,i}| = m \quad |D^+_{2,i}| = l \quad |D^-_{2,i}| = k \quad n-m > l-k}{(D^+_1, D^-_1) > (D^+_2, D^-_2)} \; \mathit{LexiPref}((D^+_1, D^-_1), (D^+_2, D^-_2), i)$$

$$2 \quad \frac{j > i \quad |D^+_{1,j}| = n \quad |D^-_{1,j}| = m \quad |D^+_{2,j}| = l \quad |D^-_{2,j}| = k \quad n-m \neq l-k}{\mathit{LexiPref}((D^+_1, D^-_1), (D^+_2, D^-_2), i) \text{ is inapplicable}} \; \mathit{LexiPref}((D^+_1, D^-_1), (D^+_2, D^-_2), i)uc$$

$$3 \quad \frac{d_1 \in D \quad \ldots \quad d_n \in D \quad \pi(d_1) = i \quad \ldots \quad \pi(d_n) = i}{|D_i| = n} \; \mathit{Count}(D, i, \{d_1, \ldots, d_n\})$$

$$4 \quad \frac{}{|D_i| = 0} \; \mathit{Count}(D, i, \varnothing)$$

$$5 \quad \frac{d_1 \in D \quad \ldots \quad d_n \in D \quad \pi(d_1) = i \quad \ldots \quad \pi(d_n) = i}{\mathit{Count}(D, i, D' \subset \{d_1, \ldots, d_n\}) \text{ is inapplicable}} \; \mathit{Count}(D, i, D')uc$$

Figure 1: Inference schemes

4.2 Inference rules for deriving preferences

The last premise of the *PrS* scheme in fact compares proposals by looking both at reasons pro and reasons con a proposal. The reasons pro are the desires that a proposal realises and the reasons con are the desires that it violates and precludes from being satisfied by any proposals considered separately elsewhere. Thus the problem of determining the last premise of the *PrS* scheme is formally identical to the problem studied in Bonnefon & Fargier [5], namely, comparing sets of positive and negative arguments for decisions. Although Bonnefon & Fargier are motivated by [1]'s argumentation-based model of multiple criteria decision making, their insights arguably equally apply to the present setting. Bonnefon & Fargier compare various combination methods on their empirical validity, that is, on how well they correspond to the preferences that people actually state. They show that one of the methods, the Lexi rule, has high empirical validity. Here we use this rule to compare sets of satisfied and violated desires.

As in [5], we assume that desires can be of varying importance. This importance is assumed to be a total preorder, which means that a set of desires can be stratified into subsets of different levels of importance. For each proposal both the set D^+ of sufficient criteria that it satisfies and the set D^- of necessary criteria that it violates are stratified into levels in this way, resulting in pairs of such sets at different levels. Then two proposals are compared by stepwise comparing their corresponding pairs of desires per level. If the proposals are equal at all levels, they are overall equal, otherwise their preference relation is determined at the highest level at which they differ in preference. The comparison per level takes place as follows: for each proposal the number of necessary criteria of that level that it violates is subtracted from the number of sufficient criteria of that level that it satisfies. The proposal with the highest resulting number is then preferred at that level.

These ideas are formalised as follows. Importance levels are expressed by integers, such that a higher integer indicates higher importance; $\pi(s) = i$ denotes that s has importance i. If D is a set of desires, then D_i is the subset of desires with importance i, i.e. $D_i = \{d \in D \mid \pi(d) = i\}$. The Lexi preference is then defined as follows:

$$(D^+_1, D^-_1) > (D^+_2, D^-_2) \Leftrightarrow \exists i \text{ such that}$$

1. $|D^+_{1,i}| - |D^-_{1,i}| > |D^+_{2,i}| - |D^-_{2,i}|$ and
2. $\forall j > i : |D^+_{1,j}| - |D^-_{1,j}| = |D^+_{2,j}| - |D^-_{2,j}|$.

In [16], an argumentation framework for deriving lexicographic preferences is presented, as a way of ordering single sets in terms of an ordering of their elements. Although thus this problem is simpler than the one studied in this paper, the inference schemes of [16] can be adapted to model [5]'s Lexi rule, so that now *pairs* of sets can be compared in terms of an ordering on their elements. This results in the inference schemes in Figure 1. Note that these schemes assume a given preference ordering on desires; by embedding these schemes in the E-ASPIC framework as defeasible inference rules, these preferences can be the outcome of an argumentation process as part of the second phase of our deliberation model.

The first scheme states that the desires satisfied and violated by proposal 1 are preferred over the desires satisfied and violated by proposal 2, if at a certain importance level i, the number of desires that 1 satisfies minus the number that it violates is higher than the number of desires that 2 satisfies minus the number that it violates.

According to the definition of Lexi preference, it is also needed that for any higher importance than i, the number of desires that 1 satisfies minus the number that it violates is equal to the number of desires that 2 satisfies minus the number that it violates. Therefore the second scheme in Figure 1 undercuts the first if this is not the case.

The counting of desires with a certain importance in a set of desires is done with the inference schemes 3 to 5, which use a kind of accrual mechanism [12]. Scheme 3 does the actual counting. Scheme 4 can be used to count 0. The last scheme is an undercutter that undercuts non-maximal counts.

Finally, it should be noted that the present inference rules are in one respect simpler than the ones in [16], namely, they assume a total instead of partial preorder . It is straightforward to generalise the present rules to partial orders along the lines of [16] but at the cost of some loss of simplicity.

4.3 The outcome of the proposal deliberation phase

During the proposal deliberation phase a second extended argumentation framework is built. At the end of the phase, all preferred

extensions are automatically identified. Note that each extension will contain at most one proposal argument, since all proposal arguments attack each other. If there is a unique extension then there is full agreement, otherwise there must be some external way to make a choice, for example, by a vote.

5. EXAMPLES

In this section we first informally discuss an example where the premises of the two criteria proposal schemes of Section 3.1 were discussed. We then present a formalised example illustrating our formal definitions.

5.1 An informal example of discussion about premises

Recall that applications of all schemes can be attacked on their premises and that our model allows for any set of argument schemes for attacking or supporting these premises, and an extensive debate may result. Thus it could be that an opponent objects to the criteria proposed to identify a desirable feature (the second premise of the *SCS* or *NCS* scheme). As an example, consider the issues relating to capital punishment found in Supreme Court cases such as *Furman v Georgia*[2]. In the various opinions there was much argument about a criterion that might be used to indicate that a punishment had the desiderata of *Effective Deterrence*. Several proposals, on both sides, were put forward, but all were flawed: the fact of the matter is that the evidence from studies investigating the the deterrent effect on the murder rate of capital punishment compared to life imprisonment is inconclusive, and arguments based on psychology also fail to secure universal support. Thus while all the Justices were agreed that *Effective Deterrence* is a desirable feature of a punishment, there was no consensus on the criterion that should be used to establish this.

Arguments concerning another proposed desiderata, *Retribution* were quite different, since here the attack was on the desirability premise. Whereas some justices, such as Burger, considered retribution a long standing and quite proper feature of punishment, opponents of capital punishment such as Brennan and, most eloquently, Marshall, argued that retribution could not be considered desirable in a free society. For example, Marshall said:

> Retaliation, vengeance, and retribution have been roundly condemned as intolerable aspirations for a government in a free society. ... If retribution alone could serve as a justification for any particular penalty, then all penalties selected by the legislature would by definition be acceptable means for designating society's moral approbation of a particular act. The "cruel and unusual" language would thus be read out of the Constitution and the fears of Patrick Henry and the other Founding Fathers would become realities.

Remember that in our setting by *desirable* is intended *desirable for society as a whole*, and therefore any dispute should be, in principle, capable of objective resolution when disputed at a particular time for a particular social group. Some goods are common to all societies at all times, such as *Public Health* and *Public Safety*. Others change in relevance over time. Whether or not, for example, *Saving Money* is desirable may depend on the current economic situation: in hard times it becomes necessary, but it is unimportant in times of prosperity. Other desiderata depend on the sort of society the society in question aspires to be: *Retribution* may be a desirable feature of punishments of societies with an emphasis on stern justice, whereas those which emphasise their civilised humanity will reject it. Note, however, that all the Justices in *Furman* do agree that while this is a proper issue for the Court to decide, it is important not to allow personal feelings to influence the issue, but that the general will be discerned. The need to decide according to the general will, setting personal preferences aside, was movingly expressed by Blackmun in his dissent upholding the death penalty:

> Cases such as these provide for me an excruciating agony of the spirit. I yield to no one in the depth of my distaste, antipathy, and, indeed, abhorrence, for the death penalty, with all its aspects of physical distress and fear and of moral judgment exercised by finite minds. That distaste is buttressed by a belief that capital punishment serves no useful purpose that can be demonstrated. For me, it violates childhood's training and life's experiences, and is not compatible with the philosophical convictions I have been able to develop. It is antagonistic to any sense of reverence for life. Were I a legislator, I would vote against the death penalty for the policy reasons argued by counsel for the respective petitioners and expressed and adopted in the several opinions filed by the Justices who vote to reverse these judgments.

On this argument while the legislators can decide to accept any desiderata they choose, the courts must respect their decisions, since their democratic election gives them a privileged position to speak as to the general will. Marshall responds to this by arguing;

> In other words, the question with which we must deal is not whether a substantial proportion of American citizens would today, if polled, opine that capital punishment is barbarously cruel, but whether they would find it to be so in the light of all information presently available.

This enables him to ignore the consistent votes of various legislatures in favour of capital punishment, and the undeniable fact that capital punishment had the support of a majority of people, and the evidence that juries were happy to authorise it, to argue that it is undesirable never the less. For Marshall, it seems, the court, since better informed than the people or the Georgia legislature, is in the best position to express the general will. This debate as to the extent to which judges can make and unmake law is never fully resolved. *Furman* perhaps represented high tide point for Marshall's view, and thereafter the need for judicial restraint became increasingly recognised.

Having given a flavour of the kinds of argument that may be deployed to debate whether criteria do indicate satisfied desires and what can be counted as desirable, we return to the argument schemes introduced in this paper, and apply them to an example.

5.2 A formal example

For our formal example we will use an issue in UK Road Traffic policy, previously used as an e-participation example in [6]. The number of fatal road accidents is an obvious cause for concern, and in the UK there are speed restrictions on various types of road, in the belief that excessive speed causes accidents. The policy issue which we will consider is how to reduce road deaths.

A number of desirable things need to be considered when deciding on such a policy.

- Our starting point is that there are an unacceptable number of deaths on the road. We are concerned with this because we desire to *Safeguard Citizens -SC*.

[2]408 U.S. 238 (1972); cf. [3]

- Road traffic accidents also cause public distress, in addition to the loss of lives. We would like to *reduce public distress - RD*

- We will suppose that we have a fixed budget. It is always desirable to *control public expenditure - PE*.

- We would also wish to see our laws, including those related to speed limits observed. This is simply a matter for *respect for the law*, whether this impacts on accidents or not - *RL*

Our police authorities always complain about a shortage of manpower, and both police and public would prefer that this scarce resource were put to solving serious crimes, rather than on traffic patrol. Therefore,

- It is desirable that *Police be available* for serious crime -*PA*

There are also some more controversial goals that we will consider here. The Civil Liberties lobby objects to the current proliferation of CCTV cameras. Even more do they object to speed cameras, which can identify the registration numbers of cars and thus effectively locate individuals. To allow us to represent their views, and because we think privacy is worth respecting, we have:

- It is desirable to *respect individual privacy -IP*

There is also a libertarian lobby, which believes that people should have maximum freedom, and sees reducing any form of regulation as desirable. Within limits we may agree:

- It is desirable to reduce regulations that restrict the freedom of individuals - *IF*.

We thus identify seven desirable features which our policy should take into consideration. We will take SC as the essential desire, since that was what our budget was allocated for. We now consider possible criteria:

C_1: *Reduction in Road Deaths* should be a sufficient criterion for *SC* since *Reduction in Road Deaths* satisfies *SC* and *SC* is desirable

C_2: *Reduction in Traffic Accidents* should be a sufficient criterion for *RD* since *Reduction in Traffic Accidents* satisfies *RD* and *RD* is desirable

C_3: *Being within Budget* should be a sufficient criterion for *PE* since *Being within Budget* satisfies *PE* and *PE* is desirable

C_4: *Reduction in Speeding Offences* should be a sufficient criterion for *RL* since *Reduction in Speeding Offences* satisfies *RL* and *RL* is desirable

N_1: *Reduction in Traffic Police* should be a necessary criterion for *PA* since *PA* requires *Reduction in Traffic Police* and *PA* is desirable

N_2: *No increase in cameras* should be a necessary criterion for *IP* since *IP* requires *No increase in cameras* and *IP* is desirable

N_3: *No additional public staff* should be a necessary criterion for *PE* since *PE* requires *No additional public staff* and *PE* is desirable

C_5: *Abolition of speed limits* should be a sufficient criterion for *IF* since *abolition of speed limits* satisfies *IF* and *IF* is desirable

There is some problem, however, with C_4. Currently many instances of speeding go unobserved and minor breaches are not prosecuted because it may appear more trouble than it is worth. With speed cameras, however, detection will improve, and prosecution will become very easy. Thus we may well expect the number of offences to rise, even though excessive, dangerous, speeding falls. None the less this is difficult to measure, and the perception will be that speeding will increase. For *RL*, therefore we will use C_4, despite our doubts.

The libertarian lobby also has an argument here: that the law will never be respected as long as it attempts to impose petty restrictions on freedom, of which speed limits are an example. Thus

N_4: *Abolition of speed limits* should be a necessary criterion for *RL* since *RL* requires *Abolition of speed limits* and *RL* is desirable

There is an incompatibility between a necessary criterion for *RL* as expressed in N_4 and a sufficient criterion as expressed in C_4, since there will be no such thing as speeding offences by which a reduction can be measured. We therefore have instances of *CN2S* and *CS2N*.

CN2S: *Reduction in Speeding Offences* does not satisfy *RL* since *RL* requires *Abolition of speed limits* and *Reduction in Speeding Offences* is not compatible with *Abolition of speed limits*.

CS2N: *RL* does not require *Abolition of speed limits* since *Reduction in Speeding Offences* satisfies *RL* and *Abolition of speed limits* is not compatible with *Reduction in Speeding Offences*.

We thus have two preferred extensions, and two alternative sets of acceptable criteria, namely,

	sufficient criteria	necessary criteria
(1):	C_1, C_2, C_3, C_4, C_5;	N_1, N_2, N_3
(2):	C_1, C_2, C_3, C_5,;	N_1, N_2, N_3, N_4

Here, however, we will resolve this by a vote and, supposing the libertarians to be in a minority, accept C_4 as a criterion for *RL* at the expense of N_4.

Next consider a set of proposals:

P_1: *Introduce More Speed Cameras*. This will, according to experience in similar countries and pilot studies reduce both accidents and fatal accidents. We expect, initially at least, there to be substantially more speeding prosecutions. We can remain within budget. Road Traffic police may be redeployed. We will, however, increase surveillance of the public and maintain speed limits.

P_2: *Increase Traffic Police*. Depending on how many additional police are used, this will either exceed budget or have insufficient impact on accidents, deaths. It does not mean more cameras but does not lead to any abolition of speed limits.

P_3: *Educate the Public*. This will exceed budget, but will reduce deaths and accidents and increase compliance without cameras although traffic police will still be required to detect those who continue to offend. Additionally, since drivers will be very aware of the dangers of speeding, it is argued that some speed limits will become unnecessary. A large number of additional staff will be required, however, to deliver the educational programme, making it very expensive.

P_4: *Abolish Speed Limits*. This fails to reduce road deaths and so does not satisfy the essential desire, and needs no further consideration, since no *PS* argument can be made for it.

We can make arguments for the three of these proposals that satisfy the essential desire.

PS_1: proposal P_1 should be adopted since
P_1 satisfies sufficient criterion C_1 for *SC* and
P_1 satisfies sufficient criterion C_2 for *RD* and
P_1 satisfies sufficient criterion C_3 for *PE*

PS_2: proposal P_2 should be adopted since
P_2 satisfies sufficient criterion C_1 for *SC* and
P_2 satisfies sufficient criterion C_2 for *RD* and
P_2 satisfies sufficient criterion C_4 for *RL*

PS_3: proposal P_3 should be adopted since
P_3 satisfies sufficient criterion C_1 for *SC* and
P_3 satisfies sufficient criterion C_2 for *RD* and
P_3 satisfies sufficient criterion C_4 for *RL* and
P_3 satisfies sufficient criterion C_5 for *IF*

We now need to consider what is preferred. We look at them pairwise. For example:

PrS: proposal P_2 is preferred over proposal P_1 since
P_1 satisfies sufficient criteria for $\{SC, RD, PE\}$
P_2 satisfies sufficient criteria for $\{SC, RD, RL\}$
P_1 violates necessary criteria for $\{IP\}$
P_2 violates necessary criteria for $\{PA\}$
$(\{\text{SC, RD,PEA}\}, \{\text{IP}\}) < (\{\text{SC,RD, RL}\}, \{\text{PA}\})$

The last premise of this argument can be verified as follows. First each of the four sets that it mentions must be stratified into levels and then we compare the two proposals per level. In detail for all three proposals:

P_1: $D_1^+ = \{SC, RD, PE\}$ $D_1^- = \{IP\}$
P_2: $D_1^+ = \{SC, RD, RL\}$ $D_1^- = \{PA\}$
P_3: $D_1^+ = \{SC, RD, RL, IF\}$ $D_1^- = \{PA, PE\}$

Let us first assume the following levels of desires (the more important, the higher the number of the level):

Level 4: $\{SC, RD, RL, PA\}$
Level 3: $\{IP\}$
Level 2: $\{IF, PE\}$
Level 1: $\varnothing$

Then the numbers at level 4 are:

P_1: $2-0=2$
P_2: $3-1=2$
P_3: $3-1=2$

So at this level all three proposals are equally preferred. Then level 3 is inspected:

P_1: $0-1=-1$
P_2: $0-0=0$
P_3: $0-0=0$

Now we know that both P_3 and P_2 are overall preferred to P_1. Next we shift to level 2 for P_2 and P_3:

P_2: $0-0=0$
P_3: $1-1=0$

So with this ordering on the set of desires P_2 and P_3 are overall equally preferred and an arbitrary choice between the two proposals has to be made.

More precisely, recall that the three instances PS_1, PS_2 and PS_3 of the *PS* scheme for the three proposals all attack each other. Now since P_1 is inferior to both P_2 and P_3, we have two instances of the *PrS* scheme that attack the attack of PS_1 on, respectively, PS_2 and PS_3, leaving the reverse attacks intact. So PS_2 and PS_3 both strictly defeat PS_1. Moreover, since P_2 and P_3 are equally preferred, we cannot instantiate the *PrS* scheme in either way for P_2 and P_3, so both attacks between PS_2 and PS_3 are left intact and these arguments defeat each other. This then results in two preferred extensions, one containing PS_2 and the other containing instead PS_3.

One way to make P_2 preferred over P_3 without affecting the outcome for P_1 is to move *IF* from level 2 to level 1. Then the score of P_3 at level 2 changes to $0-1=-1$. We can then instantiate the *PrS* scheme for the conclusion that P_2 is preferred over P_3. The result is that the attack from PS_3 on PS_2 is attacked, so that PS_2 now strictly defeats PS_3 and a unique preferred extension results, containing only P_2. Alternatively, if we were to move *IF* up to level 3, P_3 would now score 1 at that level and we could instantiate the *PrS* scheme for the conclusion that P_3 is preferred over P_2 without descending to level 2. Now the attack from PS_2 on PS_3 is attacked, so that PS_3 strictly defeats PS_2 and a unique preferred extension results, this time containing containing only P_3.

In the ordering used so far, *PE* is at the low level of 2, suggesting that times are prosperous and expenditure is possible. But if we need to place more importance on controlling public expenditure, *PE* will rise to level 3. Now P_3 scores -1 at level 3, leaving P_2 the winner at this level. Even more financial stringency, moving *PE* to level 4 would give P_1 a score of 3 at level 4, with P_2 remaining on 2, and PS_3 falling to 1. In this case, the attacks from P_2 and P_3 on P_1 would be defeated by instantiations of *PrS*, and the unique extension would contain only P_1.

Next we consider the effect of following the second method in Section 3.2, and allowing all defensible criteria to be used rather than choosing a preferred extension at the end of Phase 1. This would reinstate $N4$, and now P_1 and P_2 would violate *RL* on the basis of this criterion. Using the original stratification of values, this would leave P_3 as the clear winner at level 4.

Similarly, if we use the third method of Section 3.2, we must choose between a preferred extension endorsing P_2 and a preferred extension endorsing P_3. Note, however, that people who support P_3 may now side with the libertarian supporters of N_4. Thus N_4 is more likely to be endorsed over C_4 if the choice is made at this stage than if we had forced a decision at the end of phase 1, before the consequences in terms of proposals were known.

6. DISCUSSION AND CONCLUSION

In this paper we have proposed a formal two-phase model of democratic policy deliberation, with a clear separation between deliberation about criteria for proposals and about how proposals satisfy them. This paper offers a theoretical framework: the effectiveness and practical benefits of the proposals would require a trial with real users addressing a real problem.

Related work especially concerns formal argumentation-based models for practical reasoning and decision making, such as [1, 2, 6]. In particular, our distinction between sufficient and necessary criteria is similar to [1]'s distinction of criteria for positive and negative goals (positive goals are states to be realised while negative goals are states to be avoided). However, unlike this and similar work, our separation in two phases allows that the premises of arguments in the second phase refer to the *outcome* of the first phase. For example, some premises of the *PS* and *PrS* schemes require that criteria resulting from the first phase are acceptable. In this respect our approach is related to [8]'s modular assumption-based frameworks, in which premises of one module can refer to a consequence notion applied to another module. In future work it would be interesting to compare the pros and cons of both methods.

Another topic for future research is to extend the present model

to allow for degrees of satisfaction of desires or constraints. Then preference arguments could consider degrees of satisfaction in a more principled and realistic way than was used in section 5. One option is to use [15]'s argumentation-based influence calculus, formalised within the E-ASPIC framework.

The general idea of separating deliberations in several phases was also suggested by [9]; however, they do not provide formalisations of argument schemes or argument evaluation. Our two phases correspond the central three phases of their eight phase model: the first to their *inform* and the second to their *propose* and *consider*.

Finally, we discuss in more detail the differences from an earlier version of this work, reported in [4]. Several changes have been made which we now consider significant improvements. Criteria proposals in phase 1 are now explicitly made relative to goals, to avoid double counting in phase 2 and to give a better treatment of conflicts between criteria proposals. Such conflicts are now restricted to conflicts on how desires can and must be satisfied. Moreover, unlike in [4] we now also deal with compatibility relations between more than two criteria. Accordingly, the three conflict schemes of Section 3.1 have been completely rewritten. The new definitions make that all preference arguments on desires take place in the proposal deliberation phase, which avoids the need for carrying over preference arguments from the first to the second phase. We have also simplified the definition of the outcome of the criteria deliberation phase. Unlike in [4] there now always is a single set of acceptable criteria (in [4] called 'admissible criteria'), which simplifies the argument schemes for the second phase. Moreover, in the second phase proposals are now compared in terms of preferences on desires instead of on criteria, which seems more natural. Another addition is the inclusion of an optional set of essential desires, which must be satisfied by any proposal. Finally, we extended our formalisation of the second phase by providing inference rules for determining proposal preference in terms of preferences on individual desires. These inference rules are based on a method from the literature that is mathematically well-founded and was in empirical research found to be empirically realistic.

Acknowledgements

Wietske Visser is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs. Her research is part of the Pocket Negotiator project with grant number VICI-project 08075, and is supervised at Delft University of Technology by Koen Hindriks and Catholijn Jonker.

References

[1] L. Amgoud, J.-F. Bonnefon, and H. Prade. An argumentation-based approach to multiple criteria decision. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 05)*, number 3571 in Springer Lecture Notes in AI, pages 269–280, Berlin, 2005. Springer Verlag.

[2] K. Atkinson, T. Bench-Capon, and P. McBurney. Computational representation of persuasive argument. *Synthese*, 152:157–206, 2006.

[3] T. Bench-Capon. Towards computational modelling of Supreme Court opinions: Furman v Georgia. In K. Atkinson, editor, *Modeling Legal Cases. A Pre-Conference Workshop at the 12th International Conference on Artificial Intelligence and Law*, volume 5 of *IDT Series*, pages 77–90, Barcelona, 2009. Huygens Editorial.

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

[5] J.-F. Bonnefon and H. Fargier. Comparing sets of positive and negative arguments: Empirical assessment of seven qualitative rules. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06)*, pages 16–20, 2006.

[6] D. Cartwright and K. Atkinson. Using computational argumentation to support e-participation. *IEEE Intelligent Systems*, 24:42–52, 2009. Special Issue on Transforming E-government and E-participation through IT.

[7] P. 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.

[8] P. Dung and P. Thang. Modular argumentation for modelling legal doctrines in common law of contract. *Artificial Intelligence and Law*, 17:167–182, 2009.

[9] P. McBurney, D. Hitchcock, and S. Parsons. The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems*, 22:95–132, 2007.

[10] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173:901–934, 2009.

[11] S. Modgil and H. Prakken. Reasoning about preferences in structured extended argumentation frameworks. In P. Baroni, F. Cerutti, M. Giacomin, and G. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 347–358. IOS Press, Amsterdam etc, 2010.

[12] H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, New York, 2005. ACM Press.

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

[14] J.-J. Rousseau. The social contract. Or principles of political right, 1762. Translated by G.D.H. Cole, public domain. Available at www.constitution.org/jjr/socon.htm.

[15] T. van der Weide, F. Dignum, J.-J. Meyer, H. Prakken, and G. Vreeswijk. Arguing about preferences and decisions. In *Proceedings of the 7th International Workshop on Argumentation in Multi-Agent Systems*, pages 229–246, Toronto, 2010.

[16] W. Visser, K. Hindriks, and C. Jonker. Interest-based preference reasoning. In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence (ICAART 2011)*, 2011.

[17] D. Walton. *Practical Reasoning: Goal-driven, Knowledge-Based, Action-Guiding Argumentation*. Rowman and Littlefield, Savage, MD, 1984.

[18] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.