Integrating Dialectical and Accrual Modes of Argumentation

Sanjay Modgil, Trevor Bench-Capon [1]
Department of Computer Science, University of Liverpool

Abstract. This paper argues that accrual should be modelled in terms of reasoning about the application of preferences to sets of arguments, and shows how such reasoning can be formalised within metalevel argumentation frameworks. These frameworks adopt the same machinery and level of abstraction as Dung's argumentation framework. We thus provide a dialectical argumentation semantics that integrates accrual, and illustrate our approach by instantiating our framework with the arguments and attacks defined by an object level formalism that accommodates reasoning about priorities over sets of rules.

Keywords. Dialectical, Accrual, Metalevel, Preferences, Abstract Argumentation.

1. Introduction

Many applications of argumentation build on Dung's seminal theory [3] and its various developments. A Dung *argumentation framework* (*AF*) consists of a binary conflict based *attack* relation R on a set A of arguments. A 'dialectical calculus' is then applied to evaluate the justified and rejected arguments. Amongst developments of AFs are those that evaluate arguments only w.r.t successful attacks (*defeats*), where x defeats y only if x attacks y, and y is not stronger than x [1,2,7].

The continuing impact of Dung's theory can be attributed to its level of abstraction, and encoding of intuitive general principles of commonsense reasoning in the dialectical calculus. One defines what constitutes an argument and attack for a logic $\mathcal{L}$, so that an AF can be instantiated by the arguments and attacks defined by a theory in $\mathcal{L}$. The theory's inferences are then defined in terms of the claims of the justified arguments, as has been shown for logic programming formalisms and a number of non-monotonic logics (Dung's theory can therefore be viewed as a *dialectical semantics* for these logics).

However, this dialectical mode of argumentation fails to accommodate the intuition that the strengths of arguments may *accrue*: while an argument x claiming c is justified at the expense of arguments $y1$ and $y2$ independently claiming $\neg c$, the *combined* strength of $y1$ and $y2$ can mean that they should collectively prevail over x. Accrual may apply when evidence for and against is used to establish the truth of the matter. While in some areas it may be sensible to use Bayesian reasoning to come to an overall estimate of the probability of the hypothesis, in other cases this is not appropriate. Consider a witness testifying that P. One does not adduce some quantifiable probability of the truth of P;

[1] Corresponding Author: Sanjay Modgil, E-mail: sanjaymodgil@yahoo.co.uk.

rather one presumptively believes P. If another witness testifies the opposite, and neither witness can be discredited, then one must make a *judgement* as to who will be believed. Given several witnesses, then the witness judged to be individually the most credible may be rejected on the basis of the cumulative weight of conflicting testimony from a number of individually less credible witnesses. Accrual may also apply in decision making contexts requiring a subjective judgement or *choice*. Consider arguments supplying reasons for alternative holiday destinations. These do not force a decision, but additionally need a subjective commitment to the relative worth of the reasons they supply. It may be that the ideal destination would have good weather, food and cultural facilities. But if a paradise offering all three cannot be found, one may need to *choose* between a place with good weather and one with culture and food. One may prefer good weather to either culture or food individually, but the *combination* of the latter two may incline one towards the second possibility. We are thus interested in cases involving judgement of evidence for which a probability based treatment is not sensible, and cases requiring a choice, where a decision must be made on the basis of weighing arguments for and against. While techniques such as Multiattribute Utility Theory have been applied to such problems, they have proved problematical in practice, and fail to model actual decision-making which typically takes place in circumstances of relative ignorance. Like [11] we advocate a treatment reflecting 'quick-and-dirty' commonsense reasoning, where people reason under resource limitations and with coarse qualitative approximations to the truth.

In [11], both the *knowledge representation* (*kr*) and *inference* approaches to accrual are reviewed. In the former (e.g. [10,12]) accruals are encoded in the knowledge base, so that as well as distinct rules (and thus arguments) expressing that P is a reason for R and Q is a reason for R, there is an additional rule (and hence argument) for P *and* Q being a reason for R, and the strength of the various accruals is expressed through a priority relation on the rules. In the inference approach (e.g., [5,6,11,13]), that [11] argues has advantages over the *kr* approach, the object level inference rules permit construction of 'super-arguments' that combine individual rules that yield the same conclusion.

This paper argues for and formalises an approach to accrual that is distinct from existing approaches in two important respects. Firstly, accrual is not handled through additional arguments, whether deriving from explicit rules or from the inference mechanism. Rather, we argue that accrual is more properly effected in the (subjective) evaluation of arguments; specifically in the reasoning about and application of preferences. We thus avoid the proliferation of rules required by the *kr* approach, many of which are somewhat artificial given that their premises are entirely independent of one another. In contrast to the inference approach we respect the individuality of the accrued arguments; they continue to provide separate orthogonal reasons for the conclusions rather than a combined super-reason. Secondly, we provide an abstract integration of accrual and dialectical argumentation. We make use of the recently introduced Metalevel Argumentation Frameworks (*MAF*s) [8] to integrate argumentation based reasoning about preferences and their application, with the object level arguments being evaluated. Since *MAF*s adopt the same basic machinery of a Dung AF, we thus integrate accrual within the dialectical mode of argumentation, and therefore provide an abstract dialectical semantics for object level logical formalisms incorporating mechanisms for accrual.

In Section 2 we review background concepts. Section 3 formalises integration of accrual in *MAF*s, and relates the formalisation to [11]'s principles of accrual. In Section 4, we show how our formalism provides both a dialectical and accrual based semantics for an object level logic in which one can reason about priorities over sets of rules. We conclude with a discussion of related and future work in Section 5.

2. Background

A Dung AF is a tuple (A, R), where $R \subseteq (A \times A)$ is an attack relation on arguments A. $x \in A$ is said to be *acceptable* w.r.t. $S \subseteq A$ iff $\forall y \in A$ s.t yRx, implies $\exists z \in S$ s.t. zRy. If S is conflict free (i.e., $\forall x, y \in S$, $(x, y) \notin R$), and all arguments in S are acceptable w.r.t. S, then S is said to be an *admissible* extension. The status of arguments is then evaluated w.r.t. extensions defined under different semantics:

Definition 1 Let S be an admissible extension of (A, R).
• S is *complete* iff S contains all arguments in A which are acceptable w.r.t S; *grounded* iff S is the minimal (w.r.t. set inclusion) *complete* extension; *preferred* iff S is a maximal *complete* extension, and *stable* iff $\forall y \notin S$, $\exists x \in S$ s.t. $(x, y) \in R$
• For $s \in \{$complete, preferred, grounded, stable$\}$:
If $x \in A$ is in at least one, respectively all, s extension(s) of (A, R), then x is said to be credulously, respectively sceptically, justified under the s semantics.

For the examples in this paper, we will assume justified arguments as evaluated under the sceptical preferred semantics (although these will always coincide with the grounded semantics), and will also refer to the labelling based evaluation of arguments [9] to assist the reader's processing of the example AFs shown. A legal labelling assigns to $x \in A$: i) 1 iff $\forall y$ s.t. yRx, y = 0; ii) 0 iff $\exists y$ s.t. yRx and y = 1, and; iii) u (for undecided) iff neither i) or ii) hold. The arguments in a preferred extension are then those labelled 1 in a legal labelling with a maximal set of arguments labelled 1.

More recently, Metalevel Argumentation Frameworks (*MAF*s) [8] categorise meta-arguments according to the claims they make *about* object level arguments and their properties and relations. These meta-arguments are organised into a Dung AF whose meta-attack relation obeys constraints imposed by the claim based characterisation.

Definition 2 A MAF is a tuple $\Delta_{\mathcal{M}}$ = ($\mathcal{A}$, $\mathcal{R}$, $\mathcal{C}$, $\mathcal{L}$, $\mathcal{D}$), where $(\mathcal{A}, \mathcal{R})$ is a Dung AF, and:

- $\mathcal{L}$ consists of a countable set of constant symbols and includes the predicates: $\{$ $justified$, $defeat$, $rejected$, $preferred$ $\}$. The set $wff(\mathcal{L})$ is defined by the following BNF (x, x_i range over constant symbols)[2]:
  $\mathcal{L} : X ::= x,\ \{x_1, \ldots, x_n\} \mid justified(X) \mid rejected(X) \mid defeat(X, X') \mid preferred(X, X')$
- The claim function $\mathcal{C}$ is defined as $\mathcal{C} : \mathcal{A} \mapsto 2^{wff(\mathcal{L})}$
- $\mathcal{D}$ is a set of constrains on $\mathcal{R}$ of the form:
  if $l \in \mathcal{C}(\alpha)$ and $l' \in \mathcal{C}(\beta)$ then $(\alpha, \beta) \in \mathcal{R}$

• $\mathcal{R}$ is said to be *defined by* $\mathcal{D}$ if whenever $(\alpha, \beta) \in \mathcal{R}$ then the claims of α and β satisfy the antecedent of some constraint in $\mathcal{D}$.
• The extensions and justified arguments of $\Delta_{\mathcal{M}}$ are the extensions and justified arguments of ($\mathcal{A}$, $\mathcal{R}$).

[2] In [8] $\mathcal{L}$ also includes val, val_pref, $audience$ and wff constructed from these predicates.

Henceforth, we may use abbreviations j, r, d and p for $justified$, $rejected$, $defeat$ and $preferred$ respectively. We may also denote an argument by its claims. E.g, if $\mathcal{C}(\gamma) = \{defeat(preferred(\{a1, a2\}, \{b\}), defeat(b, a1))\ \}$, we may denote γ by $d(p(\{a1, a2\}, b), d(b, a1))$.

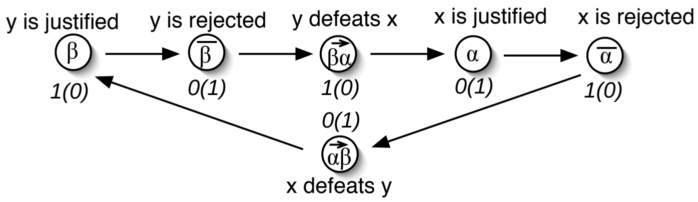

Figure 1. The MAF characterisation of a Dung AF $x \rightleftarrows y$

The basic idea of metalevel argumentation is that given an object level AF, (A, R), then the existence of an argument $x \in A$, constitutes a meta-argument $\alpha \in \mathcal{A}$ of the form 'there is an $x \in A$ that is an admissible extension of (A, R)', supporting the claim that 'x *is justified*'. The existence of an object level attack yRx, constitutes a meta-argument $\overrightarrow{\beta\alpha}$ = 'y successfully attacks x' that supports the claim 'y *defeats* x'. Since the justified status of x in the object level framework is challenged by a defeat on x, then $\overrightarrow{\beta\alpha}$ attacks α at the metalevel, and so we have the following constraint on the meta-level attack relation $\mathcal{R}$ (V, W, X, Y, Z will henceforth range over wff of $\mathcal{L}$):

$\mathbf{D1}$: if $d(Y, X) \in \mathcal{C}(\gamma)$ and $j(X) \in \mathcal{C}(\alpha)$ then $(\gamma, \alpha) \in \mathcal{R}$

y does not defeat x if y is rejected, and so $\overrightarrow{\beta\alpha}$ is attacked by a meta-argument $\overline{\beta}$ claiming 'y *is rejected*'. However, y does defeat x if y is justified, and so β claiming 'y *is justified*' attacks $\overline{\beta}$. We thus have the following metalevel constraints:

$\mathbf{D2}$: if $d(Y, X) \in \mathcal{C}(\gamma)$ and $r(Y) \in \mathcal{C}(\beta)$ then $(\beta, \gamma) \in \mathcal{R}$
$\mathbf{D3}$: if $j(X) \in \mathcal{C}(\alpha)$ and $r(X) \in \mathcal{C}(\beta)$ then $(\alpha, \beta) \in \mathcal{R}$

Fig 1 shows the MAF characterisation of a Dung AF $x \rightleftarrows y$ (together with the two labellings – the second in brackets – identifying the two preferred extensions). In [8] it is shown that:

Let Δ = (A, R), $\Delta_{\mathcal{M}}$ its MAF ($\mathcal{A}$, $\mathcal{R}$, $\mathcal{C}$, $\mathcal{L}$, $\mathcal{D}$), where $x \in A$ iff $j(x), r(y) \in \mathcal{A}$, $(y, x) \in R$ iff $d(y, x) \in \mathcal{A}$, and $\mathcal{R}$ is defined by $\{D1, D2, D3\}$. Then x is a justified argument of Δ iff $j(x)$ is a justified argument of $\Delta_{\mathcal{M}}$ (under any semantics).

Developments of AFs, including preference based AFs (PAFs) [1], value based AFs [2] and hierarchical extended *AF*s [7], are also given MAF characterisations. For example, x is a justified argument of a PAF iff $j(x)$ is a justified argument of its MAF characterisation in which object level strict preferences constitute meta-arguments that claim $preferred(x, y)$, and are constrained to meta-level ($\mathcal{R}$) attack arguments claiming $d(y, x)$. These MAF characterisations of object level argumentation allow for application of the full range of results and techniques developed for Dung *AF*s to be applied to the various object level developments of *AF*s, and provide for principled integration and extension of object level formalisms. Furthermore, in the same way that a theory's inferences can be identified by instantiating an *AF*, so MAFs can be instantiated by arguments and attacks defined by a theory, so motivating development of object level logical formalisms whose inferences can thus be identified.

3. Formalising Accrual in *MAF*s

In this section we argue that accrual is properly modelled in terms of reasoning about preferences and their undermining of the success of attacks as defeats. We formalise such reasoning within a *MAF* with metalevel constraints that explicitly obey principles of accrual identified in [11]. We thus define a dialectical argumentation semantics that integrates accrual. In the next section we instantiate such a *MAF* with the arguments and attacks defined by an object level theory, and so identify the theory's inferences as defined through a combination of dialectical and accrual modes of reasoning.

We illustrate how the dialectical mode of argumentation fails to accommodate accrual, by considering a variation on an example in [11]. Suppose an argument b claiming one should go jogging given that it is the appointed time, an argument $a1$ not to go jogging given that it is hot, and an argument $a2$ not to go jogging given that it is raining. b symmetrically attacks $a1$ and $a2$, yielding two preferred extensions $\{b\}$ and $\{a1, a2\}$; hence no argument is sceptically justified. Suppose that b is stronger than (strictly preferred to) $a1$ and stronger than $a2$. Hence the attacks from $a1$ to b and $a2$ to b do not succeed, and we are left with b asymmetrically defeating $a1$ and $a2$, so yielding the single preferred extension $\{b\}$. However, some may consider that the *combined* weight of the two independent arguments not to go jogging, outweighs b. The problem with the dialectical mode is that it considers only pair-wise relationships between arguments so that b continues to asymmetrically defeat $a1$ and $a2$, and so remains sceptically justified.

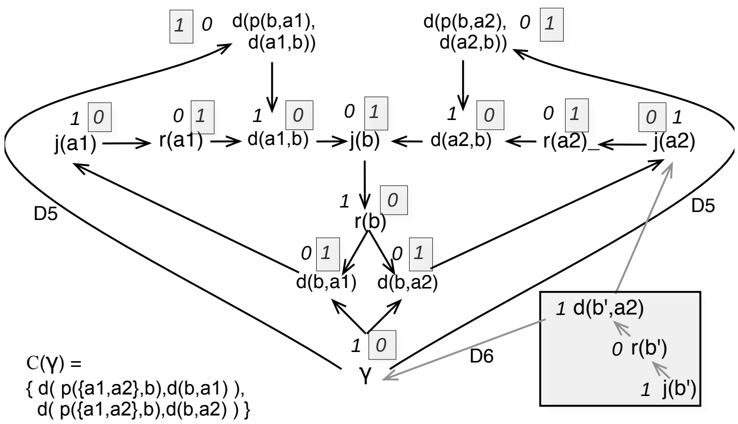

Figure 2. The accrual MAF for the jogging example with the arguments labelled as described in Section 2. The labelling shaded boxes is obtained assuming the additional arguments in the shaded box.

We claim that accrual should be modelled in terms of reasoning about preferences and their undermining of attacks. Not only do the relative strengths of individual arguments constitute reasons for undermining the success of attacks, but intuitively, the combined, or 'accrued' strengths of $a1$ and $a2$ being greater than b constitute a reason for undermining the attacks from b to $a1$ and b to $a2$. Letting upper case letters refer to accruals consisting of sets of arguments:

AC1: *The existence of a preference for accrual X over Y, based on the accrued strength of arguments in X being greater than the accrued strength of arguments in Y, is a reason for an attack from some $y \in Y$ on $x \in X$ failing to succeed as a defeat.*

We thus have meta-arguments γ with claims of the form: $defeat(preferred(X, Y), defeat(y, x)) \in \mathcal{C}(\gamma)$, where $y \in Y$, $x \in X$, and the following constraint:

D4 : if $d(p(X, Y), d(y, x)) \in \mathcal{C}(\gamma)$ and $d(y, x) \in \mathcal{C}(\beta)$ then $(\gamma, \beta) \in \mathcal{R}$

Consider the *MAF* in Figure 2 in which (apart from the argument γ that makes two claims) the arguments are denoted by the claims they make, and set brackets are omitted for singleton sets. The meta-attacks are defined by $D1 \ldots D4$ and relate meta-arguments claiming the justified and rejected status of the object level arguments b, $a1$ and $a2$, and the object level defeats between them. Given object level preferences b over $a1$, b over $a2$, and the joint preference for $a1$ and $a2$ over b, then meta-arguments claiming these preferences undermine attacks.

The preference for $\{a1, a2\}$ over $\{b\}$ preferentially undermines b's attacks on $a1$ and b's attacks on $a2$, rather than b's preference over the accrual's elements $a1$ and $a2$ undermining attacks from $a1$ to b and $a2$ to b. In general:

AC2: *Preferences defined by an accrual take precedence over the preferences defined by elements of the accrual, in that the former preferentially undermine attacks.*

The following constraint *D5* encodes **AC2** since it requires that a meta-argument γ' claiming the undermining of an attack by a preference over accruals, attacks (and so takes precedence over) any γ claiming the undermining of an attack by preferences over elements of the accruals.

D5 : if $d(p(X, Y), d(y, x)) \in \mathcal{C}(\gamma)$ and $d(p(Y', X'), d(x', y')) \in \mathcal{C}(\gamma')$ and $(X, Y) \prec_a (Y', X')$, then $(\gamma', \gamma) \in \mathcal{R}$, where:

$(X, Y) \prec_a (Y', X')$ iff $Y \subseteq Y'$, $X \subseteq X'$, and either $Y \subset Y'$ or $X \subset X'$

For example, we have attacks (labelled D5) from γ to $d(p(b, a1), d(a1, b))$ and $d(p(b, a2), d(a2, b))$ in Figure 2. Now, suppose we also had that $\{b, b1\}$ preferred to $\{a1, a2, a3\}$ (assuming additional object level argument $b1$ and $a3$). Since $(\{a1, a2\}, \{b\}) \prec_a (\{b, b1\}, \{a1, a2, a3\})$, this would attack γ and so preferentially undermine attacks from $a1$ and $a2$ to b, rather than b to $a1$ and $a2$.

Analogous to **AC2**, [11] states that: "When an accrual of arguments is applicable, that is, when there are no convincing grounds to reject the accruing elements as individual arguments, then the accrual makes its elements inapplicable." According to this principle, Prakken advocates that neither $a1$ or $a2$ are justified [3], but rather that a 'super-argument' combining $a1$ and $a2$ is justified at the expense of b. However, in our view, $a1$ and $a2$ *should* be justified. They remain individually valid reasons not to go jogging, but their acceptability in the context of a counter-argument b requires that they are jointly acceptable so that their combined weight can be taken into account. To say then, that "the accrual makes its elements inappplicable" is to refer to their applicability in an evaluative context; it is the *preferences* of the individual arguments that should not be considered applicable. Hence, Prakken's qualification – "when there are no convincing

[3] Although these arguments for the intermediate labelled versions of the conclusions *are* justified in [11].

grounds to reject the accruing elements" – on the applicability of the accrual (which he states as a separate principle: "flawed arguments may not accrue"), amounts in our view to the defeat of an element of an accrual invalidating the undermining of an attack by a preference involving that accrual. For example, if $a2$ is defeated by some b' contradicting $a2$'s premise that it is raining, then the accrued weight of $a1$ and $a2$ being greater than b should no longer preferentially undermine b's attacks on $a1$ and $a2$ since otherwise $a1$ would inappropriately be justified.

Suppose that instead we preferred $\{b\}$ to $\{a1, a2\}$, preferentially undermining attacks from $a1$ and $a2$ to b, so that b now defeats $a1$ and $a2$. We would similarly want that b''s defeat of $a2$ invalidate the use of the preference in undermining the attacks. This is because $\{a1, a2\}$ may be weaker than $a1$ alone, so that $\{b\}$ may not be preferred to $\{a1\}$. Finally, observe that we would obviously not want b's defeats of $a1$ and $a2$ to invalidate the use of the preference, since it is these defeats that are effectively decided by the preference in the first place. We thus have the following principle (analogous to [11]'s "flawed arguments may not accrue") and constraints:

AC3: *A preference for accrual X over Y undermines an attack from an argument in Y on an argument in X, if no $y \in Y$ is defeated by some $z \notin X$, and no $x \in X$ is defeated by some $z \notin Y$.*

> $\mathbf{D6}$: if $d(p(X,Y), d(y,x)) \in \mathcal{C}(\gamma)$, and $d(z,x) \in \mathcal{C}(\beta)$, $z \notin Y$, $x \in X$, then $(\beta, \gamma) \in \mathcal{R}$
>
> $\mathbf{D7}$: if $d(p(X,Y), d(y,x)) \in \mathcal{C}(\gamma)$, and $d(z,y) \in \mathcal{C}(\beta)$, $z \notin X$, $y \in Y$, then $(\beta, \gamma) \in \mathcal{R}$ [4]

For the jogging example, given the AF ($a1 \rightleftarrows b \rightleftarrows a2$) and the preferences $\{b\} > \{a1\}$, $\{b\} > \{a2\}$, $\{a1, a2\} > \{b\}$, then the justified arguments of the MAF in Figure 2 include $j(a1)$ and $j(a2)$. We also show the extra meta-arguments and attacks (shaded grey) that characterise the object level attack from b' to $a2$ (where b' contradicts $a2$'s premise that it is raining). Now b rather than $a1$ and $a2$ are justified (as indicated by the labelling in grey).

We alluded above to Prakken's third principle of accrual: "Accruals are sometimes weaker than their elements". It may be that some consider $a1$ and $a2$ to be individually stronger reasons not to go jogging, so that $a1$ and $a2$ asymmetrically defeat b and are sceptically justified. However, some may consider the combination of rain and hot to be less unpleasant, and so the accrued weight of $a1$ and $a2$ is less than b and so is preferentially a reason for undermining $a1$ and $a2$'s attacks on b, so that b now defeats $a1$ and $a2$ and b is sceptically justified. This illustrates that reasoning *about* the strengths of accruals (and more generally preferences) is itself subject to uncertainty and conflict, and so argumentation based semantics integrating accrual should accommodate object level reasoning *about* preferences. Furthermore, note that in the jogging example one could in principle prefer $\{a1, a2, b\}$ to $\{a1, a2\}$, resulting in $a1$ and $a2$ being justified. This would of course be inappropriate given that $\{a1, a2, b\}$ accrues arguments for contradictory conclusions. However, this could be precluded if one allows argumentation about preferences, so accommodating reasoned justifications for preferences over accruals. In the following definition we therefore assume a function P that maps some arguments to

[4] Note that $D6$ and $D7$ need only be applied when $|X| > 1$ and $|Y| > 1$ respectively. Space limitations preclude a detailed discussion of why this is the case.

pairwise preferences (over sets of arguments) that these arguments express (note that no commitments are made to how these preferences are defined; they may be based on unitilites, values, etc.). Hence, if $z \in A$ claims a preference for $X \subseteq A$ over $Y \subseteq A$, then we will have meta-arguments α claiming both $j(z)$ *and* $j(p(X,Y))$, and β claiming both $r(z)$ *and* $r(p(X,Y))$, where by $D3$, α will $\mathcal{R}$ attack β, and by $D2$ β will $\mathcal{R}$ attack any γ claiming $d(p(X,Y), d(y,x))$. This will be illustrated further in the following section.

Definition 3 An *Accrual MAF* (A-MAF) is a tuple $\Delta_{\mathcal{M}}$ = ($\mathcal{A}$, $\mathcal{R}$, $\mathcal{C}$, $\mathcal{L}$, $\mathcal{D}$), where $\mathcal{D}$ is the set of constraints $D1 \ldots D7$.

Let Δ be the *AF* (A, R) augmented by a partial function $P : A \mapsto 2^{\mathcal{A}} \times 2^{\mathcal{A}}$. The A-MAF $\Delta_{\mathcal{M}}$ for Δ is defined as follows:

- $\lceil x \rceil$[5] is a constant in $\mathcal{L}$ iff $x \in \mathcal{A}$
- $\mathcal{A}$ is the union of the disjoint sets $\mathcal{A}_1 \ldots \mathcal{A}_4$ where:
  1. $\alpha \in \mathcal{A}_1$, $j(\lceil z \rceil) \in \mathcal{C}(\alpha)$ iff $z \in A$, where $j(p(\lceil X \rceil, \lceil Y \rceil)) \in \mathcal{C}(\alpha)$ iff $P(z) = (X, Y)$.
  2. $\alpha \in \mathcal{A}_2$, $r(\lceil z \rceil) \in \mathcal{C}(\alpha)$ iff $z \in A$, where $r(p(\lceil X \rceil, \lceil Y \rceil)) \in \mathcal{C}(\alpha)$ iff $P(z) = (X, Y)$.
  3. $\alpha \in \mathcal{A}_3, d(x,y) \in \mathcal{C}(\alpha)$ iff $(x,y) \in R$
  4. $\alpha \in \mathcal{A}_4, d(p(X,Y), d(y,x)) \in \mathcal{C}(\alpha)$ iff $\exists z \in A$ s.t. $P(z) = (X, Y)$, and $\beta \in \mathcal{A}, d(y,x) \in \mathcal{C}(\beta)$, and $y \in Y, x \in X$.
- $\mathcal{R}$ is defined by $\mathcal{D}$.

Then: *x is a justified argument of Δ iff $j(x)$ is a justified meta-argument of $\Delta_{\mathcal{M}}$.*

Note that when an AF = (A, R) is augmented by $> \subseteq 2^{\mathcal{A}} \times 2^{\mathcal{A}}$ (rather than preferences being reasoned about in the domain of argumentation) then one can straightforwardly obtain $(A*, R*)$ where $A*$ is A augmented by arguments that map to preferences $(X, Y) \in >$, and $R*$ = $R \cup \{(z, z') | P(z) = (X, Y), P(z') = (Y, X)\}$. The A-MAF of (A, R) and $>$ would then be the A-MAF of $(A*, R*)$ and P.

To see that the constraints on an A-MAF's attack relation ensure that the principles of accrual AC1-3 are satisfied, observe that if $j(x)$ is a justified argument of $\Delta_{\mathcal{M}}$, then for every object level attack $(y, x) \in R$, $d(y, x)$ is attacked by some justified meta-argument $d(p(X,Y), d(y,x))$ and/or some justified $r(y)$. Consider the latter case[6]. For $r(y)$ to be justified there must be some justified $d(z, y)$ that attacks $j(y)$ and so ensures that $r(y)$ is reinstated against the attack by $j(y)$(see Fig.3). We can then state that the following holds (space limitations preclude inclusion of a formal proof in this paper):

Proposition 1 Let $\Delta_{\mathcal{M}}$ be the A-MAF for (A, R) augmented by a partial function P. Let $j(x)$ be a justified meta-argument of $\Delta_{\mathcal{M}}$ such that $x, y \in A$, $(y, x) \in R$. Then:

If a) $r(y)$ is not justified, or; b) it holds that: $r(y)$ is justified implies that $z = x$ for any $d(z, y)$ that is justified, then:

There are justified meta-arguments $j(p(X,Y))$ and $d(p(X,Y), d(y,x))$ such that $x \in X, y \in Y$, and:

[5] Sense quotes $\lceil\ \rceil$ are conventionally used to abbreviate metalevel representations of object level formulae.

[6] Notice that it is not necessarily the case that y is rejected ($r(y)$ is a justified meta-argument) in that y might *asymmetrically* attack x so that if x and x' are both justified, and $\{x, x'\}$ > $\{y\}$, then the asymmetric attack may be undermined and $\{x, x', y\}$ is conflict free and so possibly admissible.

1. $\forall x' \in X$, $j(x')$ is justified;
2. $\forall y' \in Y$, if $r(y')$ is justified, then any meta-argument attacking $j(y')$ (and so reinstating $r(y')$) is of the form $d(x', y)$, where $x' \in X$.
3. There is no justified $d(p(Y', X'), d(x, y))$ such that $(X, Y) \prec_a (Y', X')$ and Y', X' respectively satsify 1 and 2.

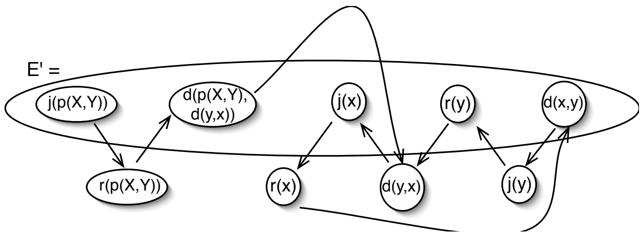

Figure 3. Meta-arguments in some $E' \subseteq E$, where E is the grounded or a preferred extension.

Informally, Proposition 1 states that if x is justified, then for any attack from y to x, if a) y is not rejected, or b) y is rejected given only that it is successfully attacked by x, then: there must be some justified argument expressing a preference for the accrual X over Y ($x \in X, y \in Y$) that undermines the attack from y to x (AC1), and: 1) All arguments in X are justified; 2) if any argument in Y is rejected, it is exclusively because of some attack originating from an argument in X (1, 2, and b equate with satisfaction of AC3); 3) there is no justified undermining of an attack from x to y by a preference for accrual Y' over an accrual X', such that X and Y are subsets of X' and Y' (AC2).

4. Instantiating A-MAFs

In this section we instantiate an A-MAF with arguments built from an object level logic that allows for reasoning about priorities over conjunctions of rules. Arguments concluding such priorities express preferences over sets of arguments, so that the instantiated A-MAF defines the inferences obtained under integration of dialectical and accrual modes of argumentation. In what follows, we assume atomic formulae built from a first order language containing the nullary predicate $pref$, and complex formulae built using the connectives $\Rightarrow$ $\neg$, $\wedge$ and $>$. We distinguish *priority formulae* of the form $X > Y$, where X and Y are conjunctions of atomic formulae.

Definition 4 A theory Γ is a set of rules $r : L_1 \wedge \ldots L_m \Rightarrow L_n$, where:
• Each unique rule name r is an atomic first order formula
• Each L_i is an atomic first order formula or a priority formula, or such a formula preceded by strong negation $\neg$.

As usual, a rule with variables is a scheme standing for all its ground instances. For any atom A or priority formula P, we say that A (P) and $\neg A$ ($\neg P$) are the complement of each other. In the metalanguage, $\overline{L}$ denotes the complement of L. Henceforth, we will refer to rules by their names, and write *head*(r) and *body*(r) to respectively denote the consequent and antecedent of the rule named r. We also assume that any Γ contains rules that ensure the priority relation $>$ is closed under transitivity, in the sense that if $r_1, r_2 \in \Gamma$, then $\exists r \in \Gamma$ s.t. *body*(r) = *body*(r_1) $\wedge$ *body*(r_2), where;

i) *head*(r_1) = $Y > X$, *head*(r_2) = $Z > Y$, head(r) = $Z > X$, or;
ii) *head*(r_1) = $Y > X$, *head*(r_2) = $\neg(Z > X)$, head(r) = $\neg(Z > Y)$, or;
iii) *head*(r_1) = $Z > Y$, *head*(r_2) = $\neg(Z > X)$, head(r) = $\neg(Y > X)$.

Definition 5 Given a theory Γ, an argument x is either:

1. a tree of rules s.t. each node $r : L_1 \wedge \ldots L_m \Rightarrow L_n$ has child nodes with rules $r_1 \ldots r_m$, where for $i = 1 \ldots m$, head(r_i) = L_i, and $r, r_1 \ldots r_m \in \Gamma$, and each leaf node of x is a rule with an empty antecedent, and no two distinct rules have the same head (so excluding arguments with circular chains of reasoning); or
2. a tree with the special root node '*pref*', each of whose child nodes is the root node r_i of a tree of type 1, where head(r_i) is a priority formula.

Table 1. A theory and its arguments and attack relation

Γ	A'
$r1 :\Rightarrow b, r2 : b \Rightarrow a$	$y1 = [r1, r2]$
$r3 :\Rightarrow c, r4 : c \Rightarrow a$	$y2 = [r3, r4]$
$r5 \Rightarrow d, r6 : d \Rightarrow \neg a$	$z1 = [r5, r6]$
$r7 :\Rightarrow e, r8 : e \Rightarrow \neg a$	$z2 = [r7, r8]$
$r9 :\Rightarrow r6 > r2$	$p1 = [r9, pref]$, $P(p1)$ = $(\{z1\}, \{y1\})$
$r10 :\Rightarrow r8 > r2$	$p2 = [r10, pref]$, $P(p2)$ = $(\{z2\}, \{y1\})$
$r11 :\Rightarrow r2 \wedge r4 > r6 \wedge r8$	$p3 = [r11, pref]$, $P(p3)$ = $(\{y1, y2\}, \{z1, z2\})$
$r12 :\Rightarrow r6 \wedge r8 > r2 \wedge r4$	$p4 = [r12, pref]$, $P(p4)$ = $(\{z1, z2\}, \{y1, y2\})$
$r13 :\Rightarrow f, r14 : f \Rightarrow r11 > r12$	$q1 = [r13, r14, pref]$, $P(q1)$ = $(\{p3\}, \{p4\})$
R' = $y1 \rightleftarrows z1, y1 \rightleftarrows z2, y2 \rightleftarrows z1, y2 \rightleftarrows z2, p3 \rightleftarrows p4$	

Definition 6 For any argument x, L is a conclusion of x iff L is the head of some rule in x. Let A be the arguments defined by Γ. For any $x, y \in A$:

x and y *attack* each other (i.e., $(x, y), (y, x) \in R$) on (L, L') iff L is a conclusion of x and L' is a conclusion of y, where L' = $\overline{L}$, or if L = $X > Y$ then L' = $Y > X$.

Arguments with root node $pref$ link together arguments concluding priorities over conjunctions of rules names. Thus, one can define pairwise preferences over sets of arguments w.r.t. priorities linked in a single $pref$ argument. Henceforth, $rules(L, Z)$ will denote $\{r | r$ is a rule in $z, z \in Z$, and $head(r) = L\}$, where Z is a set of arguments.

Definition 7 Let A be the arguments defined by Γ, and $X, Y \subseteq A$ s.t. $\{(L_1, L_1'), \ldots, (L_n, L_n')\}$ is the non-empty set of pairs of conclusions s.t. for $i = 1 \ldots n$, $\exists x \in X, y \in Y$, x and y attack on (L_i, L_i').
Let z be an argument with root node $pref$. Then $P(z)$ = (X, Y) iff for $i = 1 \ldots n$, $\bigwedge(rules(L_i, X)) > \bigwedge(rules(\overline{L_i}, Y))$ is a conclusion of z.

To enhance readability we henceforth describe propositional examples and write arguments as sequences of paths in a tree. Consider arguments $x1 = [r1 :\Rightarrow b, r2 : b \Rightarrow c]$, $x2 = [r3 :\Rightarrow c]$ and $y1 = [r4 :\Rightarrow \neg b, r5 : \neg b \Rightarrow \neg c]$ which attack each other on the pairs $(b, \neg b)$ and $(c, \neg c)$. Suppose rules $r6 :\Rightarrow r1 > r4$ and $r7 :\Rightarrow r2 \wedge r3 > r5$. Then z with root node $pref$ consists of the two paths $[r6, pref], [r7, pref]$, and $P(z) = (\{x1, x2\}, \{y1\})$.

Example 2 Consider the example Γ in Table 1 in which a subset A' and R' of the arguments and attack relation defined by Γ are shown. The instantiated A-MAF (given by Definition 3 is shown in Figure 4, in which:
$\mathcal{C}(\pi 3) = \{d(p(\{y1, y2\}, \{z1, z2\}), d(\alpha, \beta)) \mid \alpha \in \{z1, z2\}, \beta \in \{y1, y2\}\}$;
$\mathcal{C}(\pi 4) = \{d(p(\{z1, z2\}, \{y1, y2\}), d(\alpha, \beta)) \mid \alpha \in \{y1, y2\}, \beta \in \{z1, z2\}\}$.

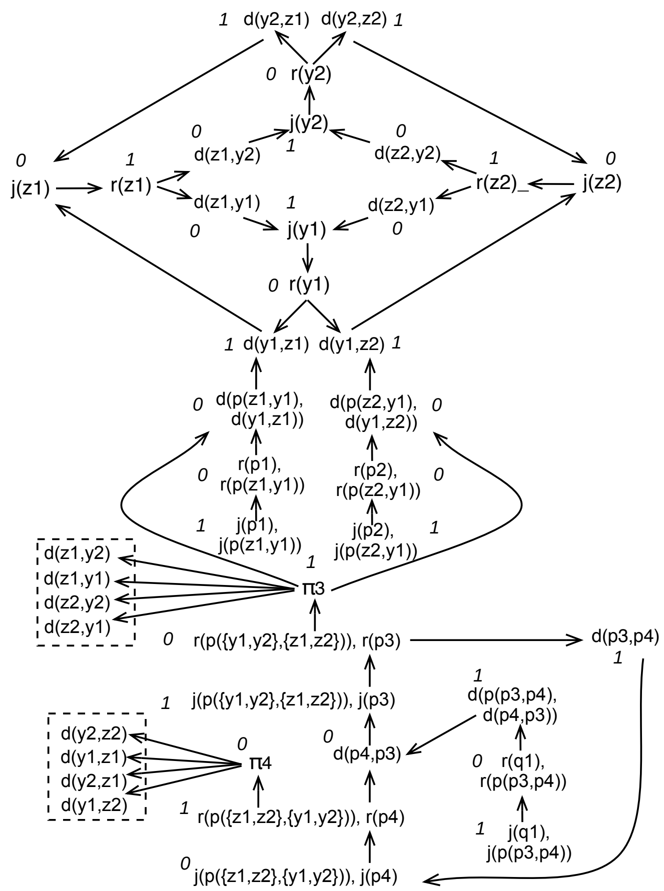

Figure 4. The A-MAF for the theory in Example 2 (to ease readability some arguments – those surrounded by dotted lines – are repeated).

In general, we say that α is an inference of Γ iff α is the conclusion of an argument x defined by Γ, and $j(x)$ is a sceptically justified argument of the A-MAF instantiated by Γ. Thus, the justified arguments of Figure 4's A-MAF (i.e., those labelled 1) identify a rather than $\neg a$ as an inference of Example 2's theory. Although arguments $z1$ and $z2$ concluding $\neg a$ are individually preferred to $y1$ concluding a, $p3$ concludes that the accrual $\{y1, y2\}$ is stronger than $\{z1, z2\}$, and $q1$ justifies a preference for this pairwise comparison over the contrary pairwise comparison concluded by $p4$.

5. Conclusions

We have argued that accrual is more properly effected through reasoning about and application of preferences in meta-argumentation frameworks that adopt the standard dialectical mode of argumentation. By contrast, existing approaches to accrual adopt either the inference or knowledge representation approach. Furthermore, they either require somewhat ad-hoc mechanisms to ensure satisfaction of accrual principles (e.g., the labelling mechanism in [11]), or formalise accrual for specific logics (e.g., [6]), or make commitments to the structure of, and interactions between arguments (e.g., [13] and [4] in which arguments are *not* evaluated using Dung's dialectical calculus), or do not accommodate dialectical argumentation (e.g., [5]). Our approach is the first to integrate accrual within Dung's dialectical theory, while preserving the theory's level of abstraction so that the inferences of various instantiating logics can be identified under integration of the dialectical and accrual modes of reasoning, and where such logics may encode reasoning about the strengths of accruals. To substantiate this claim we have shown how the inferences of a theory in such a logic are identified by the justified arguments of the theory's instantiated A-MAF. One can then apply the results and techniques developed for Dung *AF*s, to the instantiated A-MAFs. For example, argument game proof theories and algorithms for AFs [9] can be applied to A-MAFs. Future work can then investigate how efficiency gains can be obtained. For example, in an argument game, a player could in one move play arguments $d(X,Y)$ and $j(X)$, given that his counterpart will always be able to play $r(X)$ in response to $d(X,Y)$, which in turn can always be countered by $j(X)$. If played together the counterpart could then either attack $d(X,Y)$ or $j(X)$. One could thus eliminate unnecessary rounds without impacting on the game's outcome.

References

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002.

[2] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[3] 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.

[4] T. Gordon and D. Walton. Proof burdens and standards. In I. Rahwan and G. Simari, editors, *Argumentation in AI*. Springer-Verlag, 2009.

[5] P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11(1):113–131, 1995.

[6] M. J. Gómez Lucero, C. Chesñevar, and G.R. Simari. On the accrual of arguments in defeasible logic programming. In *IJCAI'09: Proc. 21st Int. joint Conf. on Artifical intelligence*, pages 804–809, 2009.

[7] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[8] S. Modgil and T.J.M Bench-Capon. Metalevel argumentation. *Technical Report*, page www.csc.liv.ac.uk/research/ techreports/techreports.html, 2009.

[9] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in AI*. Springer-Verlag, 2009.

[10] J. L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, 1995.

[11] H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *ICAIL '05: Proc. 10th Int. Conf. on Artificial intelligence and law*, pages 85–94, 2005.

[12] H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6:231–287, 1998.

[13] Bart Verheij. Accrual of arguments in defeasible argumentation. In *Proc. 2nd Dutch/German Workshop on Nonmonotonic Reasoning*, pages 217–224, 1995.