Computational Models of Argument
S. Modgil et al. (Eds.)

doi:10.3233/978-1-61499-906-5-97

Instantiating Metalevel Argumentation Frameworks

Anthony P. YOUNG [a,1], Nadin KOKCIYAN [a], Isabel SASSOON [a], Sanjay MODGIL [a], and Simon PARSONS [a]
[a] *Department of Informatics, King's College London*

Abstract. We directly instantiate metalevel argumentation frameworks (MAFs) to enable argumentation-based reasoning about information relevant to various applications. The advantage of this is that information that typically cannot be incorporated via the instantiation of object-level argumentation frameworks can now be incorporated, in particular information referencing (1) preferences over arguments, (2) the rationale for attacks, and (3) the dialectical effect of critical questions that shifts the burden of proof when posed. We achieve this by using a variant of ASPIC$^+$ and a higher-order typed language that can reference object-level formulae and arguments. We illustrate these representational advantages with a running example from clinical decision support.

Keywords. Formal models for argumentation, type theory

1. Introduction

Argumentation theory is concerned with the theory and implementation of artificial intelligence systems that perform *argumentation-based reasoning*, which can resolve conflicting information and present the reasons for or against a claim, such as recommending a decision [1,4,25]. In argumentation theory, such claims are conclusions of arguments, which are explained by being inferred from premises and rules of that argument [21], as well as showing how other arguments that disagree with such claims are not justified (e.g. [16]). *Abstract argumentation frameworks* (AFs) formalise the idea of what it means for an argument to be *justified* based on how arguments disagree with each other [11]. However, AFs by themselves do not represent and enable argumentation-based reasoning *about* all the information relevant to a given reasoning task, such as preferences. *Preference-based argumentation frameworks* (PAFs) incorporate preferences between arguments that are assumed to be given exogenously, say by the values of an audience in the case of *value-based argumentation frameworks* (VAFs), and are not themselves the objects of reasoning and disagreement [2,8,20,24]. On the other hand, *extended argumentation frameworks* (EAFs) [15,19] (with earlier work by [3,9]) have shown how one can incorporate argumentation-based reasoning about possibly conflicting preferences.

However, EAFs are limited with respect to being able to comprehensively incorporate reasoning about preferences. Even when instantiated with logical structure [19],

[1]Corresponding Author: Department of Informatics, King's College London, Bush House, Strand Campus, 30 Aldwych, WC2B 4BG, London, United Kingdom. E-mail: peter.young@kcl.ac.uk.

EAFs can only refer to when one rule within an argument is more preferred than another rule. But there may be situations where we would like to refer to when an argument *as a whole* is more preferred to another argument. For example, in clinical decision support, an argument for one course of treatment T_1 may be more preferred than another argument for another course of treatment T_2, because T_1 is elicited from clinical practice guidelines that are more relevant to the patient being treated, as compared to T_2.

In addition to reasoning about preferences, one might want to represent reasons for attacks between arguments. In logic-based argumentation, the rationale for an attack is the presence of contradictory information. However there may be other kinds of reasons for attacks. For example, in medical reasoning where only one of two drugs can be administered, e.g. due to costs or undesirable side effects, and there is no intrinsic logical contradiction precluding joint administration. *Metalevel argumentation frameworks* (MAFs) [17] can potentially represent such reasons for attacks, given that they include meta-arguments with claims of the form $attk(a,b)$, which state that the object-level argument a attacks object-level argument b. However, meta-arguments as presented in [17] have no internal structure, and serve only in providing a uniform representation of AFs, PAFs, VAFs and EAFs. In this paper, we investigate how working directly with MAFs by instantiating them with structure, can enable a richer representation of the range of possible reasons for preferences and attacks.

Arguments can be constructed through the instantiation of *argument schemes* (ASes) [27], which provide general patterns of arguments, and can be questioned with a given scheme's associated *critical questions* (CQs). Given an argument a instantiating an AS, CQs have two functions: (1) they point to possible counter-arguments to a, and (2) they question the presumptions of a, and so shift the burden of proof such that further arguments must be put forward to argue for the presumptions questioned. Until this burden of proof is met, a cannot be said to be justified. As interrogative information is not declarative, it is currently not obvious how to represent questions and their effects in logic-based instantiations of object-level argumentation frameworks. We therefore investigate how structured MAFs can allow us to represent this second interrogative effect of CQs.[2]

In this paper, we show how working directly in the metalevel (i.e. without referring to a prior object-level framework), and by instantiating MAFs with structure, can (1) enable the representation of a *wider range of reasons* behind preferences and attacks, and (2) model the effect of how asking CQs can shift the burden of proof. We endow meta-arguments with structure using *defeasible ASPIC*$^+$ – *ASPIC*^+_D [14,20] – where ASPIC^+_D arguments are now meta-arguments. This work is similar to [22], although the main difference is that we work directly in the metalevel without using support relations to anchor a prior object-level AF to the MAF. In order to refer to object-level information, we use a higher-order typed language [7] that is sufficiently expressive to refer to object-level formulae, arguments, attacks, preferences and the posing of questions. Starting in Section 2.3, we illustrate these ideas with a running example from clinical decision support.

In Section 2, we recap abstract, metalevel and structured argumentation theory, as well as questions in the context of argument schemes and critical questions. In Section 3, we make use of a higher-order typed language and provide a structured account of metalevel argumentation using this language. In Section 4 we conclude and comment on related and future work.

[2]CQs are thus an AS-specific case of a *why locution* [23]. This effect of shifting the burden of proof is more generally exemplified by why locutions in dialogical argumentation [16].

2. Background

2.1. Metalevel Argumentation

Recall from [11] that an **abstract argumentation framework** (AF) is a directed graph $\langle A, R\rangle$ where A and $R \subseteq A^2$ are, respectively, the arguments and attacks ($(a,b) \in R$, denoted $R(a,b)$, means that a attacks b). In what follows let $S \subseteq A$. We say S is **conflict-free** (cf) iff $S^2 \cap R = \varnothing$. Let $d(S) := \{a \in A \,|\, (\forall b \in A)\,[R(b,a) \Rightarrow (\exists c \in S)\,R(c,b)]\}$. We say S is **self-defending** (sd) iff $S \subseteq d(S)$. Then: S is an **admissible set** iff S is both cf and sd; S is a **complete extension** iff S is an admissible set satisfying $d(S) \subseteq S$; S is a **preferred extension** iff S is a $\subseteq$-maximal complete extension; S is a **stable extension** iff S is cf and attacks all arguments outside of it. Then S is the **grounded extension** iff S is the $\subseteq$-least complete extension. The **set of Dung semantics is** $\mathscr{S} := \{\text{complete, preferred, stable, grounded}\}$. For $s \in \mathscr{S}$ and $a \in A$, we say a is **sceptically (credulously) justified w.r.t.** s iff a is in all (some) s-extensions.

Metalevel argumentation [17] formalises the idea of how one can argue about the justification status of the arguments in AFs, PAFs, VAFs and EAFs (see Section 1). In the case of AFs, given $\langle A, R\rangle$, its corresponding **metalevel argumentation framework** (MAF) is the abstract argumentation framework $\langle A_m, R_m\rangle$, where

$$A_m := \{jus(a), rej(a) \,|\, a \in A\} \cup \{attk(a,b) \,|\, a,b \in A,\, R(a,b)\}$$

$$R_m := \{(jus(a), rej(a)), (rej(a), attk(a,b)), (attk(a,b), jus(b)) \,|\, a,b \in A,\, R(a,b)\}.$$

A_m is the **set of meta-arguments** and R_m the **set of meta-attacks** for $\langle A, R\rangle$. Here, $\langle A, R\rangle$ is the **object-level** AF. Intuitively, given $a,b \in A$, $jus(a)$ (similarly, $rej(a)$) is the meta-argument claiming that the object-level argument a is justified (rejected). The meta-argument $attk(a,b)$ is an argument accounting for the existence of the attack $R(a,b)$. The reasons for the meta-attacks are as follows. Let $a,b \in A$. The argument a is justified iff all attacks against it fail. Therefore, any attack, say from b, is a threat against the claim that a is justified, hence $attk(b,a)$ meta-attacks $jus(a)$. If b is attacking a, then b must be justified in order for the attack to be valid. Therefore, $rej(b)$ meta-attacks $attk(b,a)$. See [17, Section 3.1] for why $jus(a)$ meta-attacks $rej(a)$. Under this setup, [17, Theorem 2] states that for $s \in \mathscr{S}$, $jus(a) \in A_m$ is sceptically (credulously) justified w.r.t. s (in $\langle A_m, R_m\rangle$) iff $a \in A$ is sceptically (credulously) justified w.r.t. s (in $\langle A, R\rangle$).

MAFs are useful because they provide a uniform representation of AFs and their important generalisations (specifically PAFs, VAFs and EAFs) such that the method for calculating justified arguments follows the simpler and more intuitive method of AFs. This also opens up the way for generalisations of labelling semantics [10] and argument game proof theories [18] to these more general frameworks.

In the case of EAFs [15][17, Section 3.6], where if $R(a,b)$, then the argument c expressing the preference $a < b$ can attack the attack $R(a,b)$, and is represented by the meta-argument $aattk(c,a,b)$, which meta-attacks $attk(a,b)$, and is meta-attacked by $rej(c)$ because c must be justified in order for this attack on an attack to be valid.

Example 2.1. *Consider the EAF with arguments a,b,c, attack $R(a,b)$, and the argument c attacks the attack $R(a,b)$. The meta-arguments are $jus(a)$, $rej(a)$, $jus(b)$, $rej(b)$, $jus(c)$, $rej(c)$, $attk(a,b)$ and $aattk(c,a,b)$. These arguments and their meta-attacks are illustrated in the MAF in Figure 2.1.*

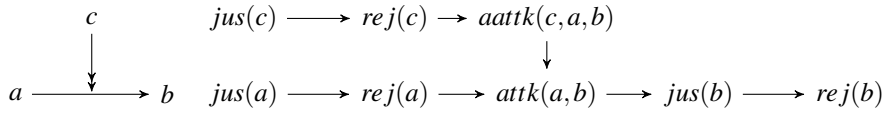

Figure 2.1. The EAF (Left) and Corresponding MAF (Right), from Example 2.1

The MAF representation of EAFs also has a correspondence result [17, Theorem 7].

2.2. *Structured Argumentation with ASPIC^+_D*

We recap the defeasible fragment of ASPIC$^+$ [5,20] – ASPIC^+_D [14] – which is less complex than the full ASPIC$^+$ and is readily implemented, while still retaining sufficient expressive power for our purposes. Let $\mathscr{L}$ be a set of well-formed formulae (wffs). Let $-:\mathscr{L}\to\mathscr{P}(\mathscr{L})$ be the **contrary function**, such that $\overline{\theta}\subseteq\mathscr{L}$ denotes the set of wffs that disagree with θ. Let $k\in\mathbb{N}$ and $(\forall 1\leq i\leq k)\,\theta_i,\phi\in\mathscr{L}$. A **defeasible rule** is denoted as $(\theta_1,\ldots,\theta_k\Rightarrow\phi)$.[3] Let $\mathscr{R}_d$ be the set of defeasible rules. The **naming function** is a partial function $n:\mathscr{R}_d\to\mathscr{L}$. The structure $\mathbf{AS}:=\langle\mathscr{L},-,\mathscr{R}_d,n\rangle$ is called an **argumentation system**.

Let $\mathscr{K}_p\subseteq\mathscr{L}$ be a distinguished subset of wffs called the **set of ordinary premises**. We call the pair $AT:=\langle\mathbf{AS},\mathscr{K}_p\rangle$ an **argumentation theory**. Given AT, we can construct arguments from $\mathscr{K}_p$ and $\mathscr{R}_d$ as follows.

1. (Base) If $\theta\in\mathscr{K}_p$, then $[\theta]$ is a **singleton argument** with **premises** $Prem([\theta]):=\{\theta\}$, **conclusion** $Conc([\theta]):=\theta$, **defeasible rules** $DR([\theta]):=\varnothing$, **subarguments** $Sub([\theta]):=\{[\theta]\}$ and **top rule** $TopRule([\theta])=*$ (i.e. undefined).

2. (Inductive) Let $k\in\mathbb{N}$ and for $1\leq i\leq k$ let A_i be an argument with premises $Prem(A_i)\subseteq\mathscr{K}_p$, conclusion $Conc(A_i)=:\theta_i\in\mathscr{L}$ and defeasible rules $DR(A_i)\subseteq\mathscr{R}_d$. Let $r:=(\theta_1,\ldots,\theta_k\Rightarrow\phi)\in\mathscr{R}_d$. Then $B:=[A_1,\ldots,A_k\Rightarrow\phi]$ is an argument with $Prem(B):=\bigcup_{i=1}^{k}Prem(A_k)\subseteq\mathscr{K}_p$, $Conc(B)=\phi\in\mathscr{L}$, $DR(B):=\{r\}\cup\bigcup_{i=1}^{k}DR(A_i)\subseteq\mathscr{R}_d$, $Sub(B):=\{A_i\}_{i=1}^{k}\cup\{B\}\subseteq\mathscr{A}$ and $TopRule(B)=r\in\mathscr{R}_d$.

Now let $\mathscr{A}$ be the set of all such arguments. Given $-$ and n, arguments $A,B\in\mathscr{A}$ can attack each other. In the following three definitions, let $a:=Conc(A)$.

1. A **undermines** B iff there is some $b\in Prem(B)$ such that $a\in\overline{b}$.
2. A **rebuts** B at $B'\in Sub(B)$ iff $b':=Conc(B')$ such that $a\in\overline{b'}$.
3. A **undercuts** B at $B'\in Sub(B)$ iff $r:=TopRule(B')$ is well-defined such that $a\in\overline{n(r)}$.

Just like ASPIC$^+$, ASPIC^+_D defines preferences $\preceq$ between arguments in terms of preferences $<$ between those arguments' defeasible rules and ordinary premises [20, Section 5]. Let $\hookrightarrow\subseteq\mathscr{A}^2$ be the relation such that $(A,B)\in\hookrightarrow$ iff A undercuts B, or [$A\not\prec B$ and (A rebuts or undermines B)], all at some appropriate $B'\in Sub(B)$. $\langle\mathscr{A},\hookrightarrow\rangle$ forms a directed graph, on which we can calculate the justified arguments and claims as in abstract argumentation.

[3] We allow for the possibility of $k=0$, in which case we have the defeasible rule $(\Rightarrow\phi)$.

2.3. Argument Schemes, Critical Questions, and a Running Example from Medicine

Argument schemes (ASes) are semi-formal representations of common patterns of everyday reasoning [27]. An **argument scheme** (AS) is thus a template that allows one to construct arguments expressed in natural or formal languages as in ASPIC$^+$, but with enough precision to make its assumptions and claim clear. For example:

Example 2.2. *Abstractly,* ***argument from expert opinion*** *is the following AS:*[4] *If (premise 1) E is an expert and (premise 2) E claims θ, then (claim) θ is true.*

Concretely, consider instantiating this in the medical domain. Eric is an overweight 52-year-old male who is suffering from high blood pressure and chronic lower back pain. In his latest visit to his doctor (general practitioner, GP), who is an expert at diagnosing strokes, she writes in his medical report[5] *that Eric had experienced a mini-stroke (transient ischemic attack, TIA). It is fair to conclude that Eric had a mini-stroke.*

ASes specialised to concrete domains can be specified [26]. The following example is an AS specialised for the medical domain.

Example 2.3. *Given (premise 1) the patient facts F, (premise 2) that the treatment goal G should be realised, and (premise 3) treatment T promotes goal G, then (claim) treatment T should be recommended. This* ***argument scheme for proposed treatment*** *[13] is a medical specialisation of Walton's sufficient condition scheme for actions [27].*

Each AS is has a set of **critical questions** (CQs), which are a means to perform due diligence when reasoning non-deductively with ASes. By questioning the truth of the assumptions of an AS, or the validity of the instantiated AS itself, CQs can either point towards counterarguments, or shift the burden of proof by requesting further reasons. If this burden of proof cannot be met, the argument based on that AS is not considered justified.

Example 2.4. *(Example 2.2 continued) Abstractly, the associated CQs that question the premises are: "Is E an expert?"; "Did E really claim θ?"; "Is θ really true?".*

As an example of the second CQ, suppose Eric questions whether his GP really wrote "TIA" on his medical report, because the handwriting really seems to read "TIN" (tubulo-interstitial nephritis, a type of kidney inflammation, which would be consistent with Eric's chronic lower back pain). Then Eric could be suspicious of whether he really did have a mini-stroke, especially as he does not remember feeling anything different on the day of the TIA. If Eric asks his GP for clarification, his question would put the burden of proof onto the GP, such that if she fails to answer the question satisfactorily, Eric could be justified in not believing that he suffered a mini-stroke.

We will now see how the effect of a CQ requesting for further reasons and so shifting the burden of proof, rather than asserting contrary information, can be represented as an attack in our structured treatment of MAFs.

[4] We present a simplified version of the full scheme that is sufficient for our purposes.

[5] Many countries, such as the United Kingdom, still use handwritten medical reports.

3. Instantiating Metalevel Argumentation Frameworks

The discussion and examples in the previous two sections motivate the following questions: how can we expand the range of reasons that can be represented in structured argumentation in relation to preferences (e.g. where we may need to refer to the argument as a whole), and attacks (e.g. in cases that are not restricted to logical contradiction, such as how performing one action excludes another)? Further, how can we represent the dialectical effect of asking CQs that shift the burden of proof? As recapped in Section 2.1, the metalevel representation of AFs and EAFs give rise to meta-arguments such as $attk(a,b)$ and $aattk(c,a,b)$, suggesting that working in the metalevel can give a rationale for attacks and preferences respectively. However, to fully account for such object level information would require for meta-arguments to have structure, i.e. that their conclusions are explained in terms of how they follow from premises via well-defined rules of inference. One way to endow MAFs with structure is to use ASPIC^+_D.

But why should these underlying arguments be meta-arguments in some MAF as opposed to just being arguments in some AF? This is because such reasoning tasks typically involve information such as preferences and the reasons for attacks, which are treated exogenously, i.e. as a "given", in standard AFs, in which case the relation $R \subseteq A^2$ is fixed. In our approach we enable reasoning about such information. Applebaum et al. have argued that working directly with MAFs can explain why arguments disagree and how preferences can nullify attacks [5], while still retaining the more straightforward approach to calculating which arguments are justified [11]. Following this reasoning, we will therefore work directly in the metalevel instead of first constructing an object-level AF and then translating into the metalevel.

We could just work in the object level and then translate to the metalevel as in [22], but the framework of [22] defines arguments and attacks in terms of what is represented in the object-level. That object-level is based on ASPIC$^+$ and suffers from the same limitations as [19] in that (e.g.) the underlying logical language cannot refer to arguments as a whole, e.g. when dealing with preferences.

Working directly in the metalevel requires that we refer to object-level arguments and their conclusions / rules (of inference) / premises with the same language. This language should at least be higher order, as conclusions are formulae and premises are sets of formulae, while arguments are not directly expressible in the language.[6] Furthermore, this language should recognise that formulae, rules and arguments are of different natures. One way of accommodating these aspects is to use a **local language**. This is a higher-order typed language that allows for the reconstruction of set theory, called a **local set theory** [7, Chapter 3], and whose models are given by elementary topoi [12]. Local languages satisfy the theorems of higher-order intuitionistic logic. The basic ideas of a local language will allow us to represent arguments, their premises, their conclusions and their rules in a uniform manner. We will see that using local languages as the underlying ASPIC^+_D language is a sufficiently flexible representational framework to offer a declarative representation of CQs by capturing the shift in burden of proof as a meta-attack; this is our answer to the question raised at the end of Section 2.3. We will continue our running example in the medical domain by building on Examples 2.3 and 2.4.

[6] As a simple example in ASPIC^+_D, if $a,b \in \mathscr{L}$ and $a \in \mathscr{K}_p$, then the rule $(a \Rightarrow b) \in \mathscr{R}_d$ is not a member of $\mathscr{L}$, and the argument $[[a] \Rightarrow b] \in \mathscr{A}$ is also not a member of $\mathscr{L}$.

3.1. Representing Reasons Behind Attacks and Preferences

We begin with ASPIC^+_D [14], where the ASPIC^+_D arguments now represent meta-arguments. We denote the argumentation system with subscript "m" meaning "meta", i.e. $\mathbf{AS}_m = \langle \mathscr{L}_m, -_m, \mathscr{R}_{d,m}, n_m \rangle$ and $AT_m = \langle \mathbf{AS}_m, \mathscr{K}_{p,m} \rangle$, where the symbols mean the same concepts as in Section 2.2. We now define $\mathscr{L}_m$ as follows.

Definition 3.1. *Our ASPIC$^+$* ***meta-language****, $\mathscr{L}_m$, has the following data:*[7] *The* ***types*** *are: the* ***truth value type*** Ω *and three* ***ground types:*** *wff for* ***well-formed formulae****, arg for* ***argument****, and rule for* ***(defeasible) rules****. Let τ be a type, then its* ***power type*** *is denoted $\mathscr{P}\tau$, which is also a well-defined type that "collects together" entities of another type; all types have power types. Further, let τ and τ' be types, then its* ***product type*** *is denoted $\tau \times \tau'$, which is also a well-defined type; all pairs of types can be combined in this way.*

We also have the following ***function symbols*** *and their* ***signatures*** *which denote the types of their inputs and outputs:* ***contrary*** $-$: *wff* $\to \mathscr{P}$*wff,* ***name*** n : *rule* $\to$ *wff,* ***premises*** *Prem* : *arg* $\to \mathscr{P}$*wff,* ***conclusion*** *Conc* : *arg* $\to$ *wff,* ***subargument*** *Sub* : *arg* $\to \mathscr{P}$*arg,* ***rule*** *DR* : *arg* $\to \mathscr{P}$*rule and* ***preference*** *Pref* : *arg* $\times$ *arg* $\to$ *wff.*

For each type we have countably many ***variables*** *available. For a type τ, we say a variable x or a term t (see below) has type τ by writing $x : \tau$ or $t : \tau$, respectively, e.g. A : arg means the variable A is of type "arg", θ : wff means the variable θ is of type "wff"... etc.*[8]

The ***terms*** *for this language $\mathscr{L}$ are as follows: a variable x of type τ, $x : \tau$, is a term, and if $t : \tau$ is a term and $f : \tau \to \tau'$ is a function symbol, then $f(t) : \tau'$ is also a term. If $\alpha : \Omega$ and $x : \tau$ are terms then the term $\{x | \alpha\} : \mathscr{P}\tau$ is well-defined.*[9] *(****Equality****) If $t : \tau$ and $t' : \tau$ are terms then $t \doteq t' : \Omega$ is also a term. (****Membership****) If $t : \tau$ and $s : \mathscr{P}\tau$ are terms, then $t \varepsilon s : \Omega$ is also a term.*[10] *We call a term of type Ω a* ***formula****.*

See [7, page 70] on how we can combine formulae with the usual constructions such as logical connectives and bounded quantifiers. For knowledge representation purposes, we define the following distinguished predicates over variables of type *arg*.

Definition 3.2. *In $\mathscr{L}_m$, we have the following distinguished predicates over variables of type arg: $jus(\cdot)$ and $rej(\cdot)$ are unary predicates, $attk(\cdot,\cdot)$ is a binary predicate, and $aattk(\cdot,\cdot,\cdot)$ is a ternary predicate. All of these predicates are of type Ω (omitted).*

Our meta-knowledge base contains information we wish to reason about, such that if we can speak of an argument, then we can speak of its components.

Definition 3.3. *Our* ***meta-knowledge base****, denoted $\mathscr{K}_{p,m}$, is a subclass of $\mathscr{L}_m$ satisfying the following argument closure condition: if A : arg $\in \mathscr{K}_{p,m}$ then $Prem(A) : \mathscr{P}$wff $\in \mathscr{K}_{p,m}$, $Conc(A)$: wff $\in \mathscr{K}_{p,m}$, $Sub(A) : \mathscr{P}$arg $\in \mathscr{K}_{p,m}$ and $DR(A) : \mathscr{P}$rule $\in \mathscr{K}_{p,m}$.*

[7] For simplicity, only a portion of the full definition of local language in [7, pp. 69 - 71] is given. This incomplete definition is sufficient for the rest of the paper, as we are interested in articulating the representational rather than the reasoning aspects of $\mathscr{L}_m$.

[8] We have $Pref(A,B)$: *wff* denote that A : *arg* is *strictly more preferred* than B : *arg*.

[9] This is a **power term**, the intuition of which is to form a new term that collects together terms of another given type. Just like in first-order logic, variables can be free or bound in $\{x | \alpha\}$. We call a term with no free variables **closed**. We can perform substitutions on free variables in the same manner as in first-order logic.

[10] Note for the syntax of the local language we use ε and $\doteq$ to be interpreted as $\in$ and $=$, respectively.

Example 3.4. *(Example 2.4 continued) Eric's GP would like to prevent future strokes by lowering Eric's blood pressure with medication. We populate $\mathscr{K}_{p,m}$ as follows. The GP may prescribe Eric one of either a low or high dose of a given drug. The arguments for these treatment options are formalised as $drug_l : arg$ and $drug_h : arg$ respectively. As prescribing one excludes the other, we also have $Conc(drug_l)\varepsilon\overline{Conc(drug_h)}_m : \Omega$ and $Conc(drug_h)\varepsilon\overline{Conc(drug_l)}_m : \Omega$. The GP prefers to prescribe the lower dose, represented as an argument $gp : arg$ such that $(\exists S \varepsilon Sub(gp))Conc(S) \doteq Pref(drug_l, drug_h) : \Omega$.*[11] *This is because the guidelines recommending the lower dose mention patient criteria that better match Eric's profile, compared to the guidelines of the higher dose.*

However, Eric argues that lifestyle changes can lower his blood pressure, represented by $ls : arg$. Furthermore, Eric argues that this would exclude the need for any medication, represented by $Conc(ls)\varepsilon\overline{Conc(drug_l)}_m : \Omega$ and $Conc(ls)\varepsilon\overline{Conc(drug_h)}_m : \Omega$. The GP disagrees with Eric's claim that lifestyle changes excludes either drug option, which we represent as $\neg Conc(ls)\,\varepsilon\overline{Conc(drug_l)}_m : \Omega$, where "$\neg$" denotes (intuitionistic) negation of formulae in $\mathscr{L}_m$.[12] *The resulting $\mathscr{K}_{p,m}$ also has all of the above terms and formulae, as well as (e.g.) $Conc(drug_h) : wff$, $Prem(ls) : \mathscr{P}wff$... etc. by Definition 3.3.*

In order to construct meta-arguments, we need suitable defeasible rules. Intuitively, these rules capture, on the basis of object-level information in $\mathscr{K}_{p,m}$, when arguments are justified, rejected, or when they attack each other.

Definition 3.5. *$\mathscr{R}_{d,m}$ contains the following **logical meta-argumentation rules**.*
*– The rules of **justification status** are $(A : arg \Rightarrow jus(A))$ and $(A : arg \Rightarrow rej(A))$.*
– Let $A : arg$, $B : arg \in \mathscr{K}_{p,m}$. We use $rb(A,B)$ to denote $Conc(A)\ \varepsilon\ \overline{Conc(B)}_m : \Omega$, $um(A,B)$ to denote $(\exists\theta\,\varepsilon\,Prem(B))Conc(A)\ \varepsilon\ \overline{\theta}_m : \Omega$, and $uc(A,B)$ to denote $(\exists r\,\varepsilon\,DR(B))Conc(A)\ \varepsilon\ \overline{n_m(r)}_m : \Omega$.
*– We then have the following three defeasible rules concerning **meta-attacks**: $(A : arg, B : arg, X(A,B) \Rightarrow attk(A,B))$ where $X \in \{rb, um, uc\}$. Further, we include three rules that are the "converses" of the meta-attack rules, which accommodate disagreement about the reasons behind the attack. They are, for $X \in \{rb, um, uc\}$: $(A : arg, B : arg, \neg X(A,B) \Rightarrow \neg attk(A,B))$. The rule of **preferences** are, for $X \in \{rb, um, uc\}$, $(A : arg, B : arg, C : arg, (\exists S \varepsilon Sub(C))Conc(S) \doteq Pref(B,A) : \Omega, X(A,B) \Rightarrow aattk(C,A,B))$.*[13]

The rules that are the "converses" of the meta-attack rules allows us to construct meta-arguments stating there is no disagreement. Generally, meta-arguments are constructed in the same manner as ASPIC$^+$, where such arguments now account for information as to the reasons behind attacks.

Example 3.6. *(Example 3.4 continued) We can construct the following meta-arguments. The justification status rules give $Jus(x) := [[x : arg] \Rightarrow jus(x)]$ and $Rej(x) := [[x : arg] \Rightarrow rej(x)]$ for $x \in \{drug_h, drug_l, ls, gp\}$, i.e. each of the four arguments has a meta-argument claiming that it is justified, and another meta-argument claiming that it is rejected. The attack rules give meta-arguments such as $Attk(ls, drug_h) := [[ls : arg], [drug_h : arg], [Conc(ls)\ \varepsilon\ \overline{Conc(drug_h)}_m : \Omega] \Rightarrow attk(ls, drug_h)]$, i.e. the con-*

[11] This is correct syntax as existential quantifiers are well-defined in the local language, see [7, page 70].
[12] For simplicity, we focus only on the case of excluding $drug_l$; the representation is analogous for $drug_h$.
[13] This captures the idea that in order to refute the attack from A to B, C either assumes or concludes that B is strictly more preferred to A, in a way that refers to the arguments A and B as a whole within the argument C.

clusion of $ls : arg$ *excludes the conclusion of* $drug_h : arg$ *(similarly with* $drug_l : arg$*), and* $\neg Attk(ls, drug_l) := [[ls : arg], [drug_l : arg], [\neg Conc(ls)\, \varepsilon\, \overline{Conc(drug_l)}_m : \Omega] \Rightarrow \neg attk(ls, drug_l)]$*. The rule of preference give the argument* $[[gp : arg], [drug_l : arg], [drug_h : arg], [(\exists S \varepsilon Sub(C))\, Conc(S) \doteq Pref(drug_l, drug_h) : \Omega], [rb(drug_l, drug_h)] \Rightarrow aattk(gp, drug_h, drug_l)] =: Aattk(gp, drug_h, drug_l)$.

The meta-attacks are defined by the meta-contrary function $-_m$ as follows.

Definition 3.7. $-_m$ *is a function on* $\mathscr{L}_m$ *that is defined only on:* $\overline{rej(A)}_m = \{jus(A)\}$, $\overline{attk(A,B)}_m = \{rej(A), aattk(C,A,B)\}$, $\overline{jus(B)}_m = \{attk(A,B)\}$, $\overline{aattk(C,A,B)}_m = \{rej(C)\}$, *and for any* $\theta : \Omega \in \mathscr{L}_m$*, we have* $\overline{\theta}_m = \{\neg\theta\}$*, and* $\overline{\neg\theta}_m = \{\theta\}$.

Example 3.8. *(Example 3.6 continued) Figure 3.1 illustrates the largest connected component of the directed graph of meta-arguments and meta-attacks.*[14]

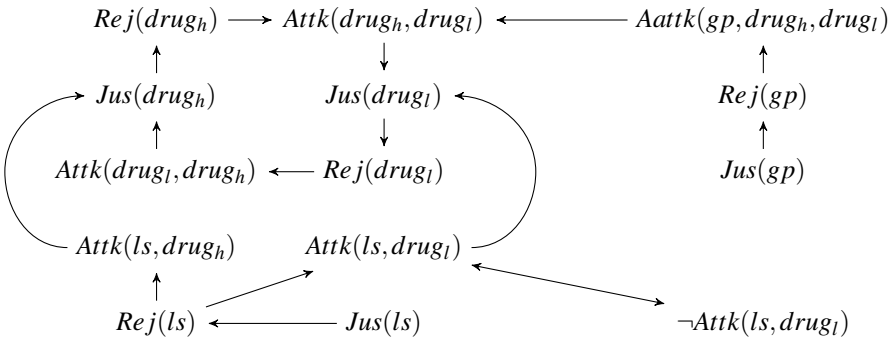

Figure 3.1. The meta-argumentation framework from Example 3.8.

Justification is calculated as in abstract argumentation [11]. It can be shown that this instantiation of ASPIC$^+$ is normatively rational [20, Section 4.2].

Example 3.9. *(Example 3.8 continued) Figure 3.1 has two stable extensions, one representing the GP's perspective by having* $\neg Attk(ls, drug_l)$ *and hence* $Jus(ls)$*,* $Jus(gp)$ *and* $Jus(drug_l)$ *justified, and the other representing Eric's perspective by having* $Attk(ls, drug_l)$*,* $Jus(ls)$ *and* $Jus(gp)$ *justified.*[15]

This example develops the example in [13], using MAFs instead of EAFs to enable a richer representation of the reasons for attacks and preferences, and accounts for multiple perspectives as extensions. Notice that one can use (meta-)preferences (which are the usual preferences of ASPIC$^+$) to arbitrate between $Attk(ls, drug_l)$ and $\neg Attk(ls, drug_l)$, and hence decide between the extensions (i.e., perspectives). However, these preferences can themselves be reasoned with in an appropriate *meta-meta-argumentation framework*. In future work, we will investigate whether this complication is necessary.

[14] There are also "stray" singleton meta-arguments such as $[Prem(ls) : \mathscr{P}wff]$ or $[DR(drug_l) : \mathscr{P}rule]$, which are always accepted. This is understood to be the meta-arguments that claim information about the object-level.

[15] In both cases, $Rej(drug_h)$ is justified. Recall that $Rej(drug_h) = [[drug_h : arg] \Rightarrow rej(drug_h)]$. This means that $[drug_h : arg]$ is a justified argument (omitted from Figure 3.1). This is not a problem as $[drug_h : arg]$ is the meta-argument claiming the *existence of* an argument for taking medication in the object level, and is not itself the argument for taking medication.

3.2. Representing the Effects of Critical Questions

Now recall from Section 2.3 that CQs can serve as either pointers to counterarguments against a given AS, or shift the burden of proof such that the proposer of the argument instantiating the AS, would need to satisfactorily answer the CQ else the argument is defeated. By adding an extra "question" type *qu* and "questioning" function symbol of signature $qu \rightarrow wff$ to $\mathscr{L}_m$, we can represent the effects of CQs as follows.

Definition 3.10. *We add a further ground type called* ***question****, denoted qu, to* $\mathscr{L}_m$*. We add a further function symbol called* ***questioning the premise****, denoted* $qp : qu \rightarrow wff$*, to* $\mathscr{L}_m$*. We add two meta-argumentation defeasible rules to* $\mathscr{R}_{d,m}$*:*

$$(A : arg, Q : qu, qp(Q)\, \varepsilon\, Prem(A) : \Omega \Rightarrow attk(Q,A)) \text{ and}$$
$$(R : arg, Q : qu, qp(Q) \eqsim Conc(R) : \Omega \Rightarrow attk(R,Q)),$$

where the first rule denotes how CQs can question an argument's premise for further reasons, and the second rule captures the effect of replying to a CQ.

Example 3.11. *(Example 2.4 continued) Suppose that Eric wants to seek clarification and asks whether his medication is needed, as it was only required because he has had a TIA, and it seems his GP wrote "TIN". We have a question by Eric* $Q : qu$ *which questions a premise of* $drug : arg$*, i.e.* $qp(Q)\, \varepsilon\, Prem(drug)$*. Therefore, we have an argument* $[[drug : arg], [Q : qu], [qp(Q)\, \varepsilon\, Prem(drug) : \Omega \Rightarrow attk(Q, drug)]]$*. This question is thus represented as a meta-argument concluding a meta-attack. However, the GP responds with the argument* $R : arg$ *which provides reasons, such as that she wrote too fast and "TIA" looks like "TIN", so* $qp(Q) \eqsim Conc(R) : \Omega$*, i.e. the premise that is being questioned by Q is addressed by the claim of the replying argument R, which provides further reasons. Therefore, we have a rebutting argument* $[[R : arg], [Q : qu], [qp(Q) \eqsim Conc(R) : \Omega] \Rightarrow attk(R, Q)]$*. This reply thus reinstates* $drug : arg$*.*[16]

In other words, the nature of the conflict here is not a conflict between contradictory pieces of information, but rather an argument should not be justified if it has not yet satisfactorily answered all questions against it. Example 3.11 is one way of showing how instantiated metalevel frameworks can provide a uniform representation of both interrogative assertions (questions) and declarative assertions (formulae and arguments).

4. Conclusions, Related Work, Future Work

We have investigated how instantiating metalevel argumentation with ASPIC^+_D and a higher-order typed language can be a model for argumentation-based decision support. By working directly with MAFs, we go beyond its role as a theoretical construct that allows for a uniform representation of abstract argumentation and its important variants. The benefits of this include the ability to reason about information that is not normally expressed in the object-level, such as the rationale for attacks, while retaining the simple and intuitive method of AFs to determine which arguments are justified.

[16] This manner of reinstating an argument by answering a question against it follows [16].

The framework articulated in this paper makes use of the theory of EAFs, MAFs, and ASPIC^+_D. Our approach differs from EAFs [15,19] because MAFs retain Dung-style reasoning rather than having to deal with reinstatement sets and collective attacks. Further, working in the metalevel can also account for attacks and questions in addition to representing preferences. This work differs from previous studies involving ASPIC^+_D [5,14] because we are using it to represent metalevel arguments and information through the use of local languages rather than representing arguments expressed in formulae of some appropriate propositional or first-order language. Some may be concerned that ASPIC^+_D may not be as expressive as ASPIC$^+$, but as [14] has argued, the defeasible fragment of ASPIC$^+$ has sufficient expressivity for many domains of application, is most likely easier to implement, and is not complicated by subtleties involving normative rationality and restricted rebuts.

To the best of our knowledge, this application of the ideas of local languages from topos theory to knowledge representation in argumentation is new. The only other work we are aware of in our field that uses ideas from topos theory is that of Atkinson et al. [6], where topos theory provides a denotational semantics for the PARMA dialogue protocol for multi-agent systems. This paper has only articulated a representational framework, and future work will aim to establish some results, by addressing, for example: (1) how can one define the distinguished predicates of Definition 3.2 in terms of more primitive predicates or are they just given for knowledge representation purposes; (2) clarifying the underlying set theory which will allow us to define $\mathscr{K}_{p,m}$ as a *subclass* of $\mathscr{L}_m$ in Definition 3.3; (3) whether the omitted features of $\mathscr{L}_m$ (Footnote 7), such as its soundness and completeness over all possible elementary topoi that interpret the language, have relevance to argumentation.

The idea of endowing metalevel argumentation with structure using ASPIC$^+$ and argumentation schemes has been considered by Müller et al. [22]. While they also consider instantiating MAFs, they retain an ASPIC$^+$ object-level framework and lift object-level information to the metalevel via support relations of *bimodal graphs*. However, we reason directly in the metalevel using a higher-order typed language. As stated in Section 3, the object-level frameworks in [22] are based on ASPIC$^+$ and cannot accommodate the representation of additional reasons for attacks and the effects of questions. Note that it could be argued that CQs are more naturally represented as being posed over the course of a dialogue (e.g. [16]). Instead, we have represented CQs in a "static" situation where the argument graph is given rather than incrementally constructed over rounds. This is not a problem because the use of CQs dynamically over the course of a dialogue would yield an AF that is evaluated statically in each round.

Future work will include a more rigorous formulation of this framework and a more thorough study of its properties, especially regarding the subtleties of $\mathscr{L}_m$ and its topos-theoretic semantics. We will also consider how domain-specific argumentation schemes, such as the schemes elicited from experts in [26], can be incorporated into our medical-style running example. More practically, we hope to investigate how such a decision support system can be implemented with a suitable argumentation engine that can reason with an appropriate representation of the input data, and relate the static evaluation of the generated frameworks with dialogical aspects of argumentation [16].

Acknowledgements: This research was supported by the UK Engineering & Physical Sciences Research Council (EPSRC) under grant #EP/P010105/1.

References

[1] L. Amgoud. Argumentation for Decision Making. In *Argumentation in Artificial Intelligence*, pages 301–320. Springer, 2009.

[2] L. Amgoud and C. Cayrol. A Reasoning Model based on the Production of Acceptable Arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[3] L. Amgoud and S. Parsons. Agent Dialogues with Conflicting Preferences. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 190–205. Springer, 2001.

[4] L. Amgoud and H. Prade. Using Arguments for Making and Explaining Decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009.

[5] A. Applebaum, Z. Li, K. Levitt, S. Parsons, J. Rowe, and E. I. Sklar. Firewall Configuration: An Application of Multiagent Metalevel Argumentation. *Argument & Computation*, 7(2-3):201–221, 2016.

[6] K. Atkinson, T. Bench-Capon, and P. McBurney. A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. *Autonomous Agents and Multi-Agent Systems*, 11(2):153–171, 2005.

[7] J. L. Bell. *Toposes and Local Set Theories: An Introduction*, volume 14 of *Oxford Logic Guides*. Oxford University Press, 1988.

[8] T. J. Bench-Capon. Persuasion in Practical Argument using Value-Based Argumentation Frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[9] G. Brewka. Dynamic Argument Systems: A Formal Model of Argumentation Processes Based on Situation Calculus. *Journal of Logic and Computation*, 11(2):257–282, 2001.

[10] M. Caminada. On the Issue of Reinstatement in Argumentation. *Logics in Artificial Intelligence*, 4160:111–123, 2006.

[11] P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and *n*-Person Games. *Artificial intelligence*, 77(2):321–357, 1995.

[12] R. Goldblatt. *Topoi: the Categorial Analysis of Logic*, volume 98. Courier Dover Publications, 2006.

[13] N. Kökciyan, I. Sassoon, A. P. Young, M. Chapman, T. Porat, M. Ashworth, V. Curcin, S. Modgil, S. Parsons, and E. Sklar. Towards an Argumentation System for Supporting Patients in Self-Managing their Chronic Conditions. In *Proceedings of the AAAI Joint Workshop on Health Intelligence (W3PHIAI 2018), to appear*, 2018.

[14] Z. Li and S. Parsons. On Argumentation with Purely Defeasible Rules. In *International Conference on Scalable Uncertainty Management*, pages 330–343. Springer, 2015.

[15] S. Modgil. Reasoning About Preferences in Argumentation Frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[16] S. Modgil. Towards a General Framework for Dialogues that Accommodate Reasoning About Preferences. In *International Workshop on Theory and Applications of Formal Argumentation*, 2017.

[17] S. Modgil and T. J. Bench-Capon. Metalevel Argumentation. *Journal of Logic and Computation*, 21(6):959–1003, 2011.

[18] S. Modgil and M. Caminada. Proof Theories and Algorithms for Abstract Argumentation Frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.

[19] S. Modgil and H. Prakken. Reasoning about Preferences in Structured Extended Argumentation Frameworks. In *COMMA*, pages 347–358, 2010.

[20] S. Modgil and H. Prakken. A General Account of Argumentation with Preferences. *Artificial Intelligence*, 195:361–397, 2013.

[21] S. Modgil, F. Toni, F. Bex, I. Bratko, C. I. Chesnevar, W. Dvořák, M. A. Falappa, X. Fan, S. A. Gaggl, A. J. García, et al. The Added Value of Argumentation. In *Agreement Technologies*, pages 357–403. Springer, 2013.

[22] J. Müller, A. Hunter, and P. Taylor. Meta-level Argumentation with Argument Schemes. In *International Conference on Scalable Uncertainty Management*, pages 92–105. Springer, 2013.

[23] H. Prakken. Coherence and Flexibility in Dialogue Games for Argumentation. *Journal of Logic and Computation*, 15(6):1009–1040, 2005.

[24] H. Prakken and G. Sartor. Argument-based Extended Logic Programming with Defeasible Priorities. *Journal of Applied Non-Classical Logics*, 7(1-2):25–75, 1997.

[25] I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*, volume 47. Springer, 2009.

[26] P. Tolchinsky, S. Modgil, K. Atkinson, P. McBurney, and U. Cortés. Deliberation Dialogues for Reasoning about Safety Critical Actions. *Autonomous Agents and Multi-Agent Systems*, 25(2):209–259, 2012.

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