Revisiting Abstract Argumentation Frameworks

Sanjay Modgil
Department of Informatics, King's College London (sanjay.modgil@kcl.ac.uk)

Abstract

This paper argues that many extensions of Dung's framework incorporating relations additional to binary attacks, are best viewed as abstractions of human rather than computational models of reasoning and debate. The paper then discusses how these additional relations may be reified into object level knowledge, thus enabling reconstruction of the extended framework as a Dung framework, and providing rational guidance for further reasoning and debate.

1 Introduction

In Dung's abstract argumentation theory [8], Dung frameworks (*DF*s) are directed graphs in which arguments (nodes) are related by binary attacks (arcs). A 'calculus of opposition' is then applied to a *DF* to determine sets of justified arguments (extensions). Dung explicitly considered the arguments and conflict based attacks as being defined, or 'instantiated', by sets of formulae (theories) in some formal logic, so that the claims of justified arguments then identify the instantiating theories' inferences. In this way, the inference relations of existing non-monotonic logics have been given argumentation based characterisations [5, 8].

Dung's abstract theory was subsequently extended in a number of directions. For example, some works formalise collective attacks from *sets* of arguments [12]. [10] included arguments that attack attacks, while [2] then generalised this idea to recursive attacks on attacks. Other works augmented *DF*s with *support* relations between arguments (e.g.,[1, 13]). While some of these works explicitly considered logical instantiations of their frameworks (e.g., [10]), many did not. This paper reviews the aforementioned extended frameworks, and then: 1) argues that they should more properly be studied as networks relating locutions as they are used in everyday reasoning and debate; 2) proposes a methodology for reconstructing these networks as Dung frameworks so as to facilitate rational reasoning and debate, and; 3) suggests ways to address the challenges that arise when obtaining these reconstructions.

The paper is organised as follows. Section 2 reviews Dung's theory and the *ASPIC*$^{+}$model of arguments and attacks [11, 15]. The latter is reviewed as reference to the internal structure of arguments will prove crucial in developing the above mentioned argument and methodology, and *ASPIC*$^+$provides a *general* account of the structure of arguments that has been shown to capture many existing approaches to argumentation. Sections 3.1 and 3.2 then review the above mentioned extensions, and argue that the additional abstract relations that many of these frameworks introduce are not warranted by logical instantiations. This is because they either fail to meaningfully abstract from underlying logical concepts, or because the interpretation of these additional abstract relations suggest that their logical instantiations can be used to reconstruct Dung frameworks. This critique then leads to the development of two lines of argument explored in Sections 3.3 and 3.4. Firstly, if the underlying logical instantiations of extended frameworks give rise to Dung frameworks that preserve the intended meaning of the additional abstract relations, then acceptability semantics defined for the extended frameworks should yield justified arguments that correspond to the justified arguments yielded by the reconstructed Dung frameworks. I show that in some cases these correspondences fail. Secondly, extended frameworks should more properly be motivated as networks that relate locutions as they are used and related in everyday reasoning and debate. These two lines of argument then lead Section 4's proposal that these networks be mapped to a computational model of structured arguments – the *ASPIC*$^+$model – and subsequently reconstructed as Dung frameworks in which the evaluated status of arguments provides feedback to users. In generating these reconstructions, one needs to 'reifiy' the abstract relations into the object level knowledge that these relations implicitly encode. However, multiple such reifications, and thus multiple reconstructed Dung frameworks, are possible. I therefore conclude by suggesting how reasoning and dialogue can be guided in order to resolve uncertainties as to what are the intended reifications. Users can be prompted to reveal the implicit knowledge encoded in the relations they assert as holding, and in so doing both enable reconstruction of Dung frameworks, and render such knowledge explicit and available for use in further reasoning and debate.

2 Background

2.1 Dung's theory of argumentation

A Dung argumentation framework (DF) is a pair $(\mathcal{A}, \mathcal{R})$, where $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation on the arguments $\mathcal{A}$. Then:

Definition 1 $S \subseteq \mathcal{A}$ is conflict free iff no two arguments in S attack each other. For any $S \subseteq \mathcal{A}$, X is acceptable w.r.t. S iff for every Y that attacks X, there is a $Z \in S$ that attacks Y (in which case Z is said to defend or 'reinstate' X). Then for any conflict free $S \subseteq \mathcal{A}$, S is an *extension* that is :

– *admissible* if every argument in S is acceptable w.r.t. S;
– *complete* if it is admissible and every argument acceptable w.r.t. S is in S;
– *preferred* if it is a maximal under set inclusion complete extension;
– *grounded* if it is the minimal under set inclusion complete extension;

– *stable* if preferred. and every argument not in S is attacked by an argument in S.

The justified arguments of $(\mathcal{A}, \mathcal{R})$ under semantics $T \in \{$preferred, grounded, stable$\}$ are those arguments in every T extension.

2.2 Arguments and attacks in the *ASPIC*$^+$framework

The remainder of this paper assumes arguments are structured as in the *ASPIC*$^+$ framework [11, 15]; i.e., as trees whose leaf nodes are premises in a given knowledge base, and whose non-leaf nodes N are either defeasible or strict inference rules of the form $\phi_1, \ldots, \phi_{n-1} \Rightarrow \phi_n$, respectively $\phi_1, \ldots, \phi_{n-1} \rightarrow \phi_n$, where for $i = 1 \ldots n-1$, N has a child node N_i that is either a premise ϕ_i, or a strict or defeasible rule with conclusion ϕ_i. Note that a premise (node) is itself an argument.

A' is then a *sub-argument* of A if A' is a sub-tree of A (including the case that A' is a leaf node (premise)). Note that A is a sub-argument of itself, whereas *proper sub-arguments* of A are sub-arguments of A excluding A itself. For simplicity I will in the remainder of this section only consider arguments with defeasible rules. Figure 1 shows four arguments B, C, D and E. Note the argument B with sub-arguments $B, B1, B2$, and $B3$.

The claim of an argument A, denoted $\texttt{Claim}(A)$, is ϕ if A's root node is a rule with consequent ϕ, or A is a single node (premise) ϕ. We also say that A forward-extends B on ϕ, equivalently B backward extends A on ϕ, if B is a proper sub-argument of A, and $\texttt{Claim}(B)$ is ϕ. Finally, $\texttt{Concs}(A)$ denotes the claims of all sub-arguments of A. For example, in Figure 1, $\texttt{Concs}(B) = \{f, b, w, q\}$.

Definition 2 A attacks B on ϕ, if $\texttt{Claim}(A)$ is the negation of some ϕ such that:
• $\phi \in \texttt{Concs}(B)$ (i.e., ϕ is a premise or consequent of a defeasible rule in B), or:
• ϕ is a name (a constant in the object level language) assigned to a defeasible inference rule in B (A is then said to 'undercut' B).

Figure 1 shows examples of attacks, from E, C and D, to B. Note that [11, 15] prohibits attacks on any ϕ that is the conclusion of a strict inference rule, since as first shown in [7], this leads to violation of rationality postulates for argumentation.

The generality of *ASPIC*$^+$accounts for this paper's assumption that arguments and attacks conform to the *ASPIC*$^+$model. One is free to choose the strict and defeasible inference rules, and the object level language in which wff ϕ are expressed. For example defeasible rules may be domain specific (akin to Reiter's default rules), such as $bird(X) \Rightarrow fly(X)$, so that given the premise $bird(tweety)$ an argument with root node $bird(tweety) \Rightarrow fly(tweety)$ claims $fly(tweety)$. Such rules may also be domain independent, e.g., *defeasible modus ponens*: $\phi, \phi \rightsquigarrow \psi \Rightarrow \psi$ ($\rightsquigarrow$ being the defeasible implication connective in the object level language). Then, given premises $bird(tweety)$, $bird(X) \rightsquigarrow fly(X)$, we have an argument claiming $fly(tweety)$, with root node: '$bird(tweety), bird(tweety) \rightsquigarrow fly(tweety) \Rightarrow fly(tweety)$' .

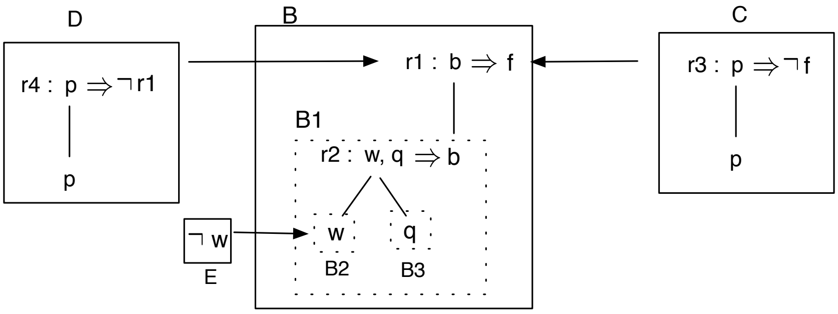

Figure 1: Propositions b, f, w, q, p respectively denote that Tweety is a *b*ird, *f*lies, has *w*ings, *q*uacks, and is a *p*enguin. r_i, $i = 1 \ldots 4$, are propositions naming rules.

Note that [11, 15] also generalise the notion of negation allowing one to specify that a wff is a contrary of another wff ($\neg$ is then a special case, i.e., ϕ is a contrary of ψ whenever ψ is of the form $\neg\phi$ or ϕ is of the form $\neg\psi$). It is this notion of contrary that [11, 15] refer to when defining attacks. [11, 15] then show that many logical instantiations of Dung frameworks and other general structured approaches to argumentation can be formalised as instances of the *ASPIC*$^+$framework, in the sense that the arguments and attacks they define are special cases of *ASPIC*$^+$arguments and attacks. For example, classical logic instantiations of Dung frameworks, where premises may be taken from a knowledge base of classical wff, and arguments are constructed using only strict classical inference rules (e.g., modus ponens etc).

3 Abstract Argumentation Frameworks: Acceptability Semantics and Instantiations

This section reviews examples of abstract argumentation frameworks (AAFs) that extend DFs with support and variants of binary attack relations. I will assume that, as in the case of DFs, these AAFs are instantiated by underlying logical theories. I then argue that in cases where abstract level relations are meaningful abstractions of underlying logical relations, one can reconstruct DFs from the underlying theories. I then conclude that : 1) the reconstructions shed light on how evaluation of the justified arguments in the AAFs may need to be modified; 2) AAFs should more properly be viewed as modelling human reasoning and debate, rather than as abstractions of underlying theories in some formal logic.

3.1 Support Relations

I begin by considering frameworks with support relations. [1]'s bipolar argumentation framework (*BAF*) is of the form $(\mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{supp})$, where $\mathcal{R}_{supp}$ is a support relation and $\mathcal{R}_{att}$ an attack relation ([1] call $\mathcal{R}_{att}$ a 'defeat relation',). The question arises as to what these support relations abstract from, in the sense that if A attacks B on ϕ, then the attack abstracts from the object level logical relationship of negation relating $\texttt{Claim}(A)$ and ϕ. [1] explicitly answer this question for specific kinds of arguments of the form (H, h) where H is a set of consistent classical wff (premises) that minimally (under set inclusion) classically entail h. Then (H, h) supports (H', h') if $h \in H'$ or $h = h'$. Generalising this notion to *ASPIC*$^+$arguments:

A supports B on ϕ if $\texttt{Claim}(A) = \phi, \phi \in \texttt{Concs}(B)$. **S1**

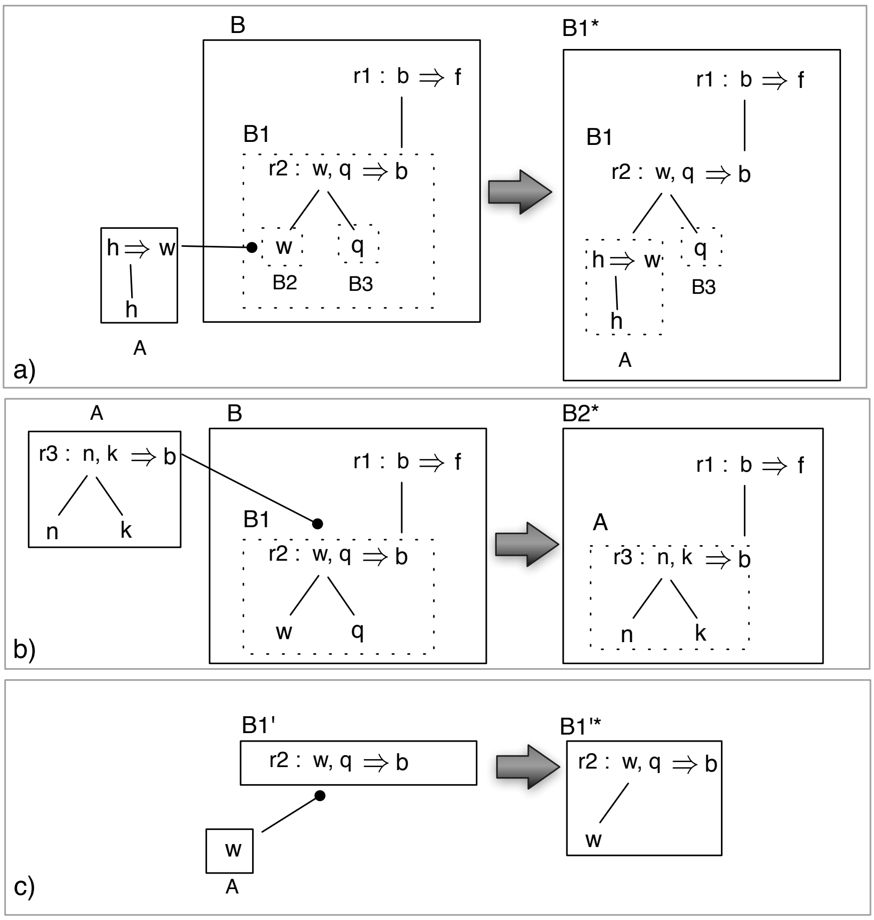

Figure 2: Support relations are represented as lines with swollen ends. Note that n, k, h respectively denote that Tweety builds *n*ests, has a bea*k*, and is feat*h*ered.

Figures 2-a) and 2-b) respectively show an *ASPIC*$^+$argument A supporting B on w and b. That **S1** is the intended interpretation of support is further testified to by the motivating example dialogues in [1], e.g., in Example 6 in [1], F = 'I concerns a problem of public health, so I is important information' supports A = 'I is important information, so we must publish it'. However, this interpretation of support then implies that if X supports Y on ϕ, then X backward extends Y on ϕ to define another argument Y^*. This is illustrated in Figures 2-a) and 2-b) : A backwards extends B on w and b respectively, so that one can 'reconstruct' arguments $B1^*$ and $B2^*$. In other words, given the same logical information, one can instantiate a *DF* consisting only of arguments and binary attacks.

Consider another example of support relations in [14], in which the 'argument' X = "The bridge should be built where slow water exists without mud (i.e.at x,y)" is said to be supported by the argument B = "Our historic survey says that slow water exists at coordinates x,y". Firstly, note that X is a rule rather than an argument, with consequent "The bridge should be built at x,y" and antecedent "slow water exists without mud at x,y". Then B supports X in the sense that X extended with the premise B, on its antecedent, yields an argument. This suggest a second distinct notion of support :

A supports B on ϕ if $\texttt{Claim}(A) = \phi$, ϕ is in the antecedent of a rule in B. **S2**

Figure 2-c) illustrates S2-support: A supports $B1'$ on w, so that one can reconstruct $B1'^*$. The example shows that s2-support does not always licence the reconstruction of arguments from the underlying logical information; $B1'^*$ is not an argument, rather it is a rule in need of a supporting argument (for q). On the other hand, in [14]'s example above, B fully supports X, enabling reconstruction of an argument.

3.2 Attack Relations

[12] and [4] extend *DF*s with collective attacks. In particular, in [12], individual arguments can be attacked by non-empty sets of arguments:

Definition 3 A Dung framework with collective attacks (AF_c) is a tuple $(\mathcal{A}, \mathcal{R}_c)$ where $\mathcal{A}$ is a set of arguments, and $\mathcal{R}_c \subseteq (2^{\mathcal{A}} \setminus \emptyset) \times \mathcal{A}$. Then:

- $S \subseteq \mathcal{A}$ is conflict free iff $\neg\exists S' \subseteq S, X \in S$ such that $(S', X) \in \mathcal{R}_c$.
- X is acceptable w.r.t. $S \subseteq \mathcal{A}$ iff $\forall \mathcal{A}' \subseteq \mathcal{A}$ such that $(\mathcal{A}', X) \in \mathcal{R}_c$, $\exists S' \subseteq S$ such that $(S', Y) \in \mathcal{R}_c$ for some $Y \in \mathcal{A}'$.

The extensions of an AF_c are then defined as in Definition 1. [12]'s motivating example considers arguments $A1$ = *Joe does not like Jack* and $A2$ = *There is a nail in Jack's antique coffee table* collectively attacking B = *Joe has no arms, so Joe cannot use a hammer, so Joe did not strike a nail into Jack's antique coffee table*. Quoting from [12], $A1$ and $A2$ "jointly provide a case for the conclusion that Joe has a struck a nail into Jack's antique coffee table". This implies that the collective attack is an abstraction of a rule relating the claims of $A1$ and $A2$ to the negation of the claim of B. This suggests we can reconstruct a DF, given that $A1$ and $A2$ can be extended with a rule '*If Joe does not like Jack* and *there is a nail in Jack's*

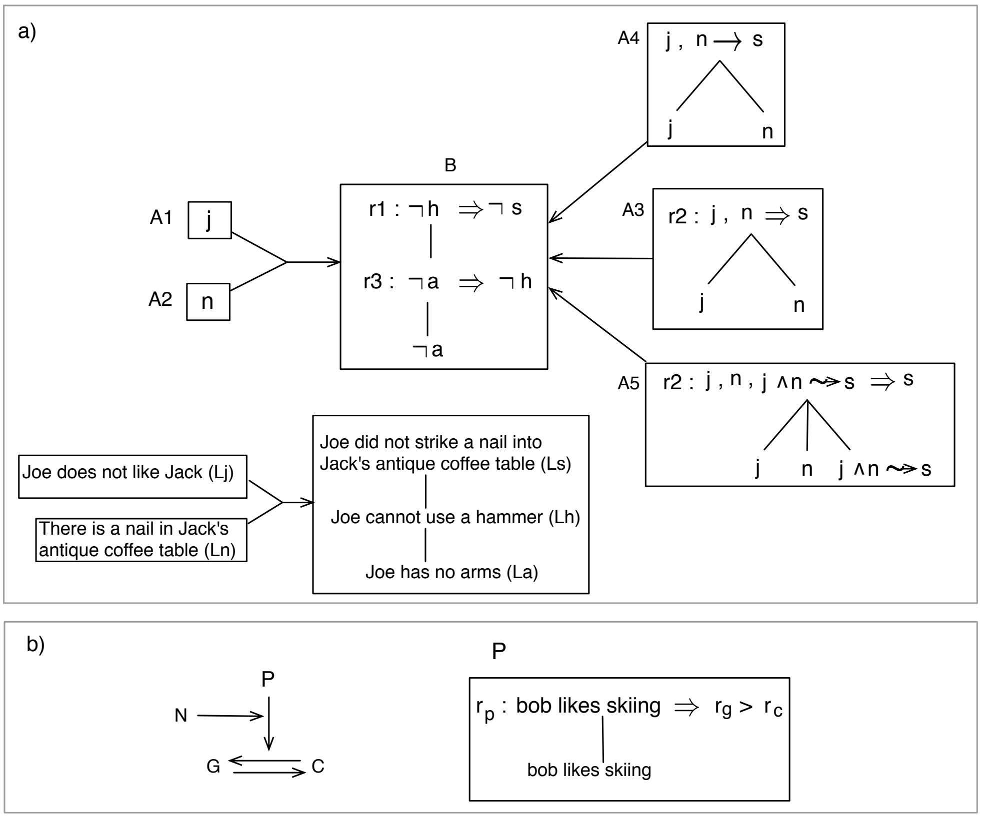

Figure 3: a) Natural language representation of locutions Lj, Ln, La, Lh, Ls involved in a collective attack are shown, with the corresponding logical formulation of the collective attack from $\{A1, A2\}$ to B, and three possible reconstructed arguments $A3$, $A4$ and $A5$ based on the collective attack. b) shows a preference attack on an attack (P), and a recursive attack on an attack (N)

antique coffee table then *Joe has a struck a nail into Jack's antique coffee table*' to define an argument A that directly attacks B on its conclusion. In general:

Definition 4 Let $X_1, \ldots, X_n$ collectively attack Y on ϕ ($\phi \in \texttt{Concs}(Y)$ or ϕ names a defeasible inference rule in Y). Then:
$recon_c(\{X_1, \ldots, X_n\}, \phi)$ is the argument X, whose root node is the defeasible rule $\texttt{Claim}(X_1), \ldots, \texttt{Claim}(X_n) \Rightarrow \neg\phi$, backward extended by arguments $X_1, \ldots, X_n$, such that X attacks Y on ϕ.

Figure 3-a) illustrates, showing how $A3$ = $recon_c(\{A1, A2\}, \neg s)$ is backward extended by $A1$ and $A2$. $A3$ then attacks B on $\neg s$.

While [12] acknowledge that collectively attacking arguments can be extended to single arguments which then attack their target, they maintain that collective attacks are still warranted by logical instantiations, since (referring Figure 3-a) it may be that $A1$ or $A2$ are attacked, but $A3$ is not attacked. But a structured account of argumentation shows this cannot be the case. Recalling Definition 2, an argument is attacked if any of its sub-arguments are attacked, so that if $A1$ or $A2$ are attacked then $A3$ is attacked (on its sub-arguments $A1$ or $A2$).

Now note that for frameworks with support relations, the logical information yielding a BAF will yield exactly one reconstructed DF. However, the logical information yielding an AF_c may yield many DFs. This is because the same collective attack may abstract from different underlying logical instantiations. For example, we have thus far ignored arguments containing strict inference rules, but $recon_c(\{X_1, \ldots, X_n\}, \phi)$ might be defined to yield arguments with top nodes $\texttt{Claim}(X_1), \ldots, \texttt{Claim}(X_n) \rightarrow \neg\phi$ (argument $A4$ in Figure 3-a)) or arguments with additional premises together with domain independent inference rules ($A5$ in Figure 3-a)). I comment further on this issue in Section 4.

A number of works extend DFs with attacks on attacks. In [10]'s Extended Argumentation Frameworks (*EAF*s), an argument P claiming a preference for G over its attacker C, attacks the attack from C to G, so that the success of C's attack on G is denied, and G is justified (Fig.3-b). For example, G ='Bob want to go to Gstaad since there is a last minute offer for Gstaad' symmetrically attacks C = 'Bob want to go to Cuba since there is a last minute offer for Cuba'. Then P expresses Bob's preference for G over C given that Bob likes skiing and so prefers ski resorts.[10] explicitly studied logical instantiations of *EAF*s, where arguments expressing preferences are instantiated by premises and rules concluding priorities over rules in the arguments over which the preferences are claimed. For example, P might be an argument with the premise 'Bob likes skiing' and defeasible rule r_p concluding that the rule r_g in G has greater priority than the rule r_c in C.

[2] then generalised *EAF*s to recursive attacks on attacks. For the example in Figure 3-b), [2] suggest that N = 'there have been no snowfalls in Gstaad for a month so it is not possible to ski in Gstaad' attacks the preference attack from P. However, unlike P, it is difficult to conceive of a logical instantiation yielding an argument N that claims a preference for an *attack* $C \rightarrow G$ over the *argument* P. Indeed, one might intuitively consider N as claiming $\neg r_p$, so undercutting P on its rule, since not being able to ski in Gstaad denies the defeasible inference step from Bob likes skiing, to a preference for (the rule in) G over (the rule in) C. Finally, note that while recursive attacks do not seem well motivated from a logical instantiation perspective, I will in Section 3.4 suggest an alternative motivation.

3.3 Acceptability Semantics for Abstract Argumentation Frameworks

The previous section's discussion suggests that if AAFs such as bipolar frameworks (BAFs) and frameworks with collective attacks (AF_cs) can be reconstructed as Dung frameworks, then one would expect a correspondence between the status of arguments in the AAFs and their status in the reconstructed DFs.

Firstly, consider that if A supports B, and B symmetrically attacks C, then the preferred extensions defined in [1] are $\{A, B\}$ and $\{C\}$, since [1] suggest that since A supports B and B attacks C then there is a supported attack from A to C, and so $\{A, C\}$ is not conflict free. Suppose now we have the *ASPIC*$^+$arguments[1] A = $[p; p \Rightarrow q]$, B = $[q; q \Rightarrow t]$, C = $[\neg t]$, where A supports B on the premise q. Then we can reconstruct the additional argument B^* =$[p; p \Rightarrow q; q \Rightarrow t]$ which also symmetrically attacks C. The preferred extensions of the reconstructed DF are $\{A, B, B^*\}$ and $\{C, A\}$ (ignoring arguments $[p]$ and $[q]$ which are irrelevant to the analysis). Thus the expected correspondence does not hold, since A is justified in in the reconstructed DF, but A is not justified in the original BAF. The discrepancy arises because it seems that in the abstract BAF, A is assumed to support B on its claim, in which case a correspondence would then hold, since in the reconstructed DF, A would symmetrically attack C. This illustrates that evaluation of arguments in a BAF needs to account for the structure of arguments and targets of support relations (ie., the sub-arguments that are supported).

There may also be a discrepancy between the justified arguments of an AF_c and its reconstructed DF. Recall that any given AF_c may yield more than one reconstruction, as a collective attack may be an abstraction of a number of different logical instantiations. In Fig.3, $\{A1, A2\}$ asymmetrically attacks B, and since no arguments attack $A1$ or $A2$ then B is not in an admissible extension of the AF_c. But if we reconstruct with the argument $A3$, then since $A3$ and B have contradictory claims obtained by application of defeasible rules, they symmetrically attack (by Definition 2), and so B can defend itself and is in an admissible extension. Furthermore, since B attacks $A3$, then in the corresponding AF_c, one would expect that B attacks $\{A1, A2\}$, but attacks on sets of arguments are not allowed in [12]. If instead we assume the reconstructed argument $A4$ then we would then have that $A4$ asymmetrically attacks B (recall from Section 2.2 that attacks cannot target the conclusions of strict rules), and so the desired correspondence would obtain.

To conclude, I have argued that acceptability semantics for AAFs need to account for the structure of arguments, such that a correspondence obtains with the acceptability semantics of the associated reconstructed DFs.

3.4 Abstract Locution Networks

In Sections 3.1 and 3.2, I argued that relations additional to binary attacks are not well motivated under the assumption that AAFs are instantiated by logical theories. In what follows I argue that they are more properly motivated under the assumption that they are required to model the way humans reason and debate.

As stated in Section 1, non-monotonic inference relations can be characterised in terms of the claims of justified arguments. This testifies to the generality of the reinstatement principle (Definition 1); a principle that is both intuitive and familiar in human modes of reasoning and dialogue. Dung's theory thus abstractly characterises both human and logic-based reasoning in the presence of uncertainty and

[1]Henceforth *ASPIC*$^+$arguments may be represented as square brackets enclosing premises and rules separated by semi-colons.

conflict, in terms of the dialectical use of attacking and defending arguments, so that argumentation-based characterisations of computational and human reasoning and dialogue can inform and enhance each other. To facilitate this bridging role requires development of argumentation models that accommodate human reasoning and dialogue as conducted in practice. This suggests a more constructive reformulation of the critique that AAFs do not adequately motivate abstract concepts and relations additional to binary attacks, in terms of formal logical instantiations. Rather, AAFs should be studied under the assumption that they are motivated by requirements for modelling relations between locutions as used in every day reasoning and debate. This is of course implied by the works reviewed in this paper, which make use of motivating and illustrative examples of every day dialogue and reasoning. [14]'s example in Section 3.1 illustrates the use of support to account for locutions that do not always consist of fully formed arguments, but may instead be rules, so that arguments are implicitly constructed piecemeal by possibly different interlocutors supplying different elements of an argument. More generally, humans clearly make statements in support of other statements, as witnessed by numerous natural language examples in [1, 13, 14] and other works utilising support. Furthermore, [12] explicitly motivate collective attacks for modelling human dialogue, giving examples of locutions submitted by different interlocutors, that combine to define collective attacks. However, I suggest that this motivation for AAFs is under-appreciated by the research community, and will in what follows suggest further implied research directions.

To begin with, I propose that AAFs should be viewed as special cases of *Abstract Locution Networks* (ALNs), in which the nodes are locutions related by binary attacks, support relations, collective attacks, attacks on attacks, recursive attacks e.t.c. Note that abstract dialectical frameworks (*ADF*s) [6] might be considered as a candidate formalism for such networks, but the technical machinery associated with *ADF*'s suggest that they are unsuitable candidates for modelling reasoning and dialogue as conducted in practice [2]. In order to now motivate future research directions, consider a software tool for single users or users engaged in dialogue, that enables: 1) entry of locutions that can in turn be linked to other locutions so as to structure rules and arguments; 2) linking of individual locutions, rules and arguments to denote relationships of support and various kinds of attack.

4 From Abstract Locution Networks to Computational Knowledge

The above described 'ALN tool' would contribute to the plethora of existing argument visualisation and mapping tools [9]. A key research goal is to then map the arguments diagrammed in these tools to computational models of argument, so that they can can be evaluated under Dung's various semantics [3], and thus inform

[2]Also note that Section 3.3's argument that acceptability semantics for *AAF*'s needs to account for the structure of arguments, also applies to the acceptability semantics defined for *ADF*s.

reasoning and debate by: ensuring that the assessment of arguments is formally and rationally grounded; enabling humans to track the status of arguments so that they can be guided in which arguments to respond to.

Assume an ALN tool and a mapping of the contained linked locutions to *ASPIC*$^+$premises, rules and arguments[3]. The key point to note is that unlike the previous section's AAFs, that are assumed to be instantiated by formal logical theories, the diagrammed ALNs are authored by humans, and the goal is to map these to computational knowledge[4] so that one can instantiate a Dung framework in order to provide dialectical feedback to users. The challenge is to then account for the fact that ALNs do not consist exclusively of arguments related by binary attacks. Section 3 then suggests a methodology for addressing this challenge.

To illustrate, consider the natural language diagramming of the collective attack in Figure 3, mapped to the *ASPIC*$^+$arguments $A1, A2, B$ and their constituent premises and rules. In order to then reconstruct a DF based on this computational knowledge, a choice has to made as to how to *reify*[5] the collective attack so as to yield the additional computational knowledge – either $j, n \rightarrow s$ or $j, n \Rightarrow s$ or $j \wedge n \rightsquigarrow s$ - that would then be used to construct either of the arguments $A3$, $A4$ and $A5$ respectively. As discussed in Section 3.3, the choice of reification and thus additional constructed arguments, will affect the evaluated status of arguments in the reconstructed DF. In other words, given the diagramming of the collective attack, there remains some uncertainty as to how to reify this attack. Indeed, such uncertainty is likely to be the norm, given that not all relevant information is explicitly articulated in everyday reasoning and dialogue; much is left implicit. In this example, not only is the additional rule needed to reconstruct the argument not rendered explicit in the locutions related by the attack, *but also the target of the attack is implicit*. How is one to disambiguate whether the locutions Lj and Ln collectively attack on La. Lh or Ls ? (although it is assumed that the attack is on Ls; hence the assumed reifications of rules concluding s). How then is one to resolve such uncertainties, so that one can deterministically reconstruct a DF in order to provide dialectical feedback ?

This issue also arises when considering *binary* support and attack relations, given the commonplace use of *enthymemes* (arguments in which information is omitted) in everyday discourse. Consider the following dialogue:

> Paul argues that "Tony Blair is no longer a public figure, the information about his affair is not in the public interest, and the information is private, so the information should not be published" (X). Trevor counter-argues with "but Blair is UN envoy for the Middle East" (Y).

---

[3]Such mappings are described (via intermediate translation to the Argumentation Interchange Format) in [3]

[4]Recall that we are not committing to a particular computational model, but any of the broad range of models shown to be instances of *ASPIC*$^+$.

[5]Note that in the previous section we referred to abstract relations such as attack, collective attacks and supports as 'abstractions' of underlying logical relations. However, given an ALN with diagrammed abstract relations, the task is to now reify these to yield computational knowledge

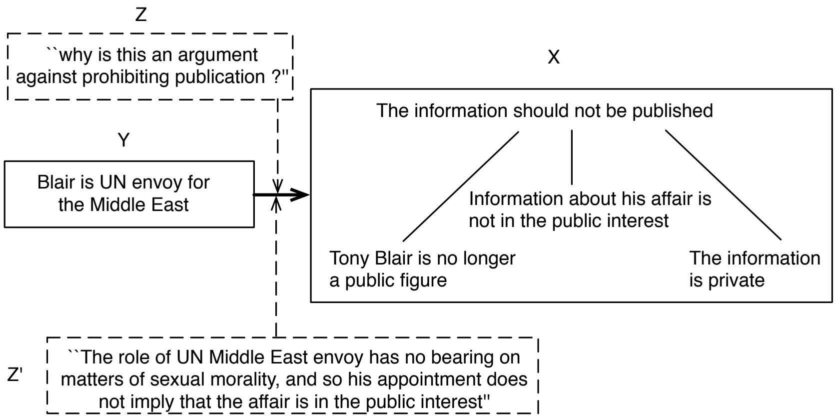

Figure 4: An attack by an enthymeme Y on X.

Y is just such an enthymeme. The very fact that Y is moved as an attack on X, but the attack is not explicitly targeted, is indicative of an incomplete rule of the form 'if someone is a UN envoy for the Middle East then (s)he is ...', where the missing information is some claim negating an element in X. We thus need to reify a *binary attack* to obtain a rule that can be used together with the premise *Blair is UN envoy for the Middle East* to yield an argument Y^* that attacks X. But then should the reified rule be $mi \Rightarrow pf$ or $mi \Rightarrow pi$ or $mi \Rightarrow \neg pr$ or $mi \Rightarrow \neg pub$, or $mi \Rightarrow \neg r_1$, where mi = *Blair is UN Middle East Envoy*, pf, pi, pr and pub respectively denote *Blair is a public figure*, *the information is in the public interest*, *the information is private*, and *the information should be published*, and r_1 names the defeasible rule in the computational representation of X ?

Notice that the same issue arises with support relations. Suppose Paul supports his argument X with B = "Blair holds no public office". Once again, we see a requirement for reifying the support relation, to obtain a rule that would then augment the premise (enthymeme) B, so yielding an argument B^* claiming $\neg pf$ or $\neg pi$ or pr or $\neg pub$ that would then s1-support X on one of its conclusions.

To summarise, in order to reconstruct a DF on the basis of locutions related in an ALN, and thus provide dialectical feedback to users, one needs to resolve uncertainties as to how the abstract relations relating locutions are reified. One way to resolve such uncertainties is through the prompting of further dialogue moves, such that responses to these moves furnish the required information to decide upon a reification. For example, suppose the dialogue above now continues as follows:

> Paul then counters by asking "why is this an argument against prohibiting publication ?" (Z). Trevor responds with "because his appointment as UN Middle East envoy implies that the information about his affair is in the public interest" (V)

Hence, the uncertainty is resolved in favour of the reification of the attack being

$mi \Rightarrow pi$. Thus, we see a form of dialectical feedback whereby further dialogical moves are prompted, resulting in rendering explicit, knowledge (V) implicitly encoded in the attack, so that this knowledge is available for use in further reasoning and debate, and can be used to reconstruct the DF to provide evaluative feedback.

I conclude by noting that attacks as conceived in Dung's theory play two roles. That Y attacks X is an abstraction of the declarative incompatibility of Y's claim and some element in X, as well as an abstract characterisation of the dialectical, procedural use of Y as a counter-argument to X. Definition 1's notion of a conflict free set accounts for the declarative denotation, whereas the notion of acceptability of arguments accounts for the dialectical denotation. Preference attacks in *EAF*s invalidate the dialectical use of attacks. However, one cannot question, in a formal logical context, the declarative basis of an attack from Y to X on ϕ, since to do so would be to question the fundamental logical principle that a formula (i.e., $\texttt{Claim}(Y)$) and its negation (i.e., ϕ) are in conflict.

However, since attacks and support relations in ALNs may implicitly encode object level knowledge, this suggests one can attack the declarative rationale for an attack (and indeed support) relation. This also suggests a motivation for [2]'s recursive attacks, which in Section 3.2 were claimed to be not well motivated by logical instantiations. Such 'rationale attacks' can shift the burden of proof to the proposer of the attacked attack, to furnish the latter's declarative rationale, e.g., Paul submits the rationale attack Z on the attack from Y to X, and Trevor then fulfils his burden of proof by providing the rationale V. Of course, Paul may (perhaps mistakenly) assume from the outset the intended rationale for the attack from Y to X, and submit an alternative rationale attack Z' on $Y \rightarrow X$: Z' = "The role of UN Middle East envoy has no bearing on matters of sexual morality, and so his appointment does not imply that the affair is in the public interest".

5 Conclusions

This paper has argued that various extensions of Dung's abstract framework should be studied under the assumption that they model human reasoning and debate, and should therefore account for the fact that locutions do not consist of fully formed arguments that can be related by binary attacks, but rather as statements, rules and incomplete arguments organised into networks in which they are related to each other in more complex ways. I then proposed reconstruction of these networks as Dung frameworks in order that reasoning and debate can be informed by rational models of argument. Such reconstruction requires reification of these relations to the object level knowledge they implicitly encode. Given that locutions often consist of incomplete arguments (enthymemes) I illustrated requirements for reification of binary attacks in addition to other relations that augment these attacks in extended frameworks. For any given relation, many such reifications are possible, with the choice of reification impacting on the evaluation of arguments in the reconstructed Dung framework that it defines. I then suggested that resolution of these choices should prompt dialogical moves that elicit replies confirming the intended reification, thus resolving the choice of reconstructed framework, and making available implicit knowledge for use in further reasoning and debate.

This paper lays foundations for a programme of research. The first task, currently underway, is to formalise reconstruction of networks of locutions as Dung frameworks consisting of *ASPIC*$^+$arguments and attacks, building on the methodology suggested in this paper. This would involve broadening the range of networks considered in this paper, to include (for example) networks in which arguments or statements are asserted as being for, or against a claim. That the latter reconstructions are possible is attested to by a recent translation of the Carneades model of argumentation to *ASPIC*$^+$[16] (Carneades models arguments for and against claims). In this paper I have also suggested how recursive attacks on attacks, and indeed attacks on supports, can intuitively be motivated in the context of everyday reasoning and debate. The second task is to then augment existing models of dialogue so that the required reconstruction of underlying Dung frameworks prompts the submission of dialogue moves for eliciting implicit knowledge. Note also, that the focus in this paper has been on 'assertive' locutions that commit speakers to the truth of expressed propositions. However, as illustrated by Section 4's dialogue in which Paul issues a 'why' locution, other types of locution will need to be considered. In this case the relationship with Trevor's locution can still be interpreted as an attack, but other types of locution and dialogues in the Walton and Krabbe typology [17], will warrant a broader range of relations considered in this paper, with different interpretations that may or may not admit reification to object level knowledge. Finally, these two tasks would contribute to the long term aim of linking tools for mapping reasoning and debate to computational models of argument and dialogue, so that the latter can rationally guide the former.

Acknowledgements; I would like to thank the anonymous reviewers for their helpful comments.

References

[1] L. Amgoud, C. Cayrol, M. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093, 2008.

[2] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. Afra: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52(1):19 – 37, 2011.

[3] F. Bex, S. Modgil, H. Prakken, and C. Reed. On logical specifications of the argument interchange format. *Journal of Logic and Computation*, doi: 10.1093/logcom/exs033, 2012.

[4] A. Bochman. Collective argumentation and disjunctive programming. *Journal of Logic and Computation*, 13 (3):405–428, 2003.

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

[6] G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Proc. of KR 2010*, pages 102–111, 2010.

[7] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[8] 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):321–358, 1995.

[9] P. A. Kirschner, S. J. Buckingham Shum, and C. S. Carr, editors. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer, London, UK, 2003.

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

[11] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195(0):361 – 397, 2013.

[12] S. H. Nielsen and S. Parsons. A generalization of dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proc. 3rd Int. Workshop on Argumentation in Multi-agent Systems*, pages 54–73, 2006.

[13] N. Oren and T. J. Norman. Semantics for evidence-based argumentation. In *In Computational Models of Argument: Proceedings of COMMA 2008*, pages 276–284, 2008.

[14] N. Oren, T. J. Norman, and A. Preece. Evidential reasoning in bipolar argumentation frameworks. In *Proceedings of the Fourth International Workshop on Argumentation in Multi-agent Systems*, 2007.

[15] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[16] B. van Gijzel and H. Prakken. Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation. *Argument and Computation*, 1:21–47, 2012.

[17] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Series in Logic and Language. State University of New York Press, Albany, NY, USA, 1995.