Weakest Link in Formal Argumentation: Lookahead and Principle-based Analysis

Chen Chen[1 [0009−0008−9775−6782]], Pere Pardo[2 [0000−0003−0181−4658]], Leendert van der Torre[1,2 [0000−0003−4330−3717]], and Liuwen Yu[2,3 [0000−0002−7200−6001]]

[1] Zhejiang University, China
[2] University of Luxembourg, Luxembourg
[3] University of Bologna, Italy

Abstract. In this paper, we introduce a new definition of weakest link attack relation assignment based on lookahead, and compare this new lookahead definition with two existing ones in the literature using a principle-based analysis. We adopt a formal framework for such attack relation assignments that was introduced by Dung in 2016. We show that our lookahead definition does not satisfy context independence, we introduce a new principle called weak context independence, and we show that lookahead weakest link satisfies weak context independence. We also show that lookahead weakest link is the closest approximation to Brewka's prioritised default logic PDL, also known as the greedy approach. For PDL, we prove an impossibility result under Dung's axioms. Our results generalise earlier findings restricted to total orders to the more general case of modular orders.

Keywords: Prioritised structured argumentation · Weakest link · Principle-based analysis · Formal argumentation · Knowledge representation and reasoning.

1 Introduction

The saga of weakest link is one of the great stories of defeasible argumentation. The idea that a chain of reasoning is as strong as its weakest link was used by John Pollock in 1995 as a way to compare the strength of arguments [30]. In Pollock's words: the strength of each conclusion is the minimum of the strengths of the inference with which it was derived and of the premises or intermediate conclusions from which it was derived [30, p. 99]. Pollock wrote a series of influential articles on defeasible reasoning that laid the foundations of formal argumentation [28–32]. Also in 1995, Dung published a seminal paper on abstract argumentation that became as well part of the foundations of formal argumentation [11]. It has been used as a general framework for instantiating (prioritised) default logic [11, 34] and defeasible logic [20], among other non-monotonic systems. These logics can be formalised in structured argumentation (e.g. ASPIC+) to generate abstract argumentation frameworks.[4] In ASPIC+, the

[4] Structured argumentation builds arguments from the rules and facts of a knowledge base. Abstract argumentation just assumes an attack relation to define sets of arguments that are collectively acceptable, while ignoring the underlying logic that defines attacks as logical conflicts.

attack relation is defined by a notion of argument strength based on weakest link or last link [23, 24].

Whether one agrees or not with Modgil and Prakken that *weakest link* is appropriate for epistemic scenarios while *last link* suits better normative scenarios [24], this choice has an impact on queries to knowledge bases and normative systems: *Do fitness-loving Scots like whisky? Should snoring professors get access to the library?* [23, 24]:

$$\text{The fitness-lover Scot} \quad \left\{ \begin{array}{c} bornInScotland \Rightarrow scottish \\ scottish \Rightarrow likesWhisky \\ fitnessLover \Rightarrow \neg likesWhisky \end{array} \right\} \qquad \text{Snoring professor at library} \quad \left\{ \begin{array}{c} snores \Rightarrow misbehaves \\ misbehaves \Rightarrow accessDenied \\ professor \Rightarrow \neg accessDenied \end{array} \right\}$$

Pollock's work and the distinction between weakest and last link in particular played a central role in formal models of structured argumentation. This important distinction between weakest and last link necessitates in fact the possibility of representing default rules —compare e.g. with Assumption-Based Argumentation (ABA) [4] or classical logic-based argumentation [3]. Principle-based analyses [8, 12, 13, 16, 19] have recently studied general properties of attack relations under various approaches to structured argumentation. However, given the long history of weakest link, it may come as a surprise that there have been few developments characterising how it can be used to instantiate abstract argumentation frameworks that capture a given logic. Starting with traditional weakest link [23, 24, 30], and the variant called disjoint weakest link [34], we explain this saga and its relation to prioritised default logic (PDL) [5] using three benchmark examples and study the following research question: how to axiomatize the attack relations that correspond to each variant of weakest link?

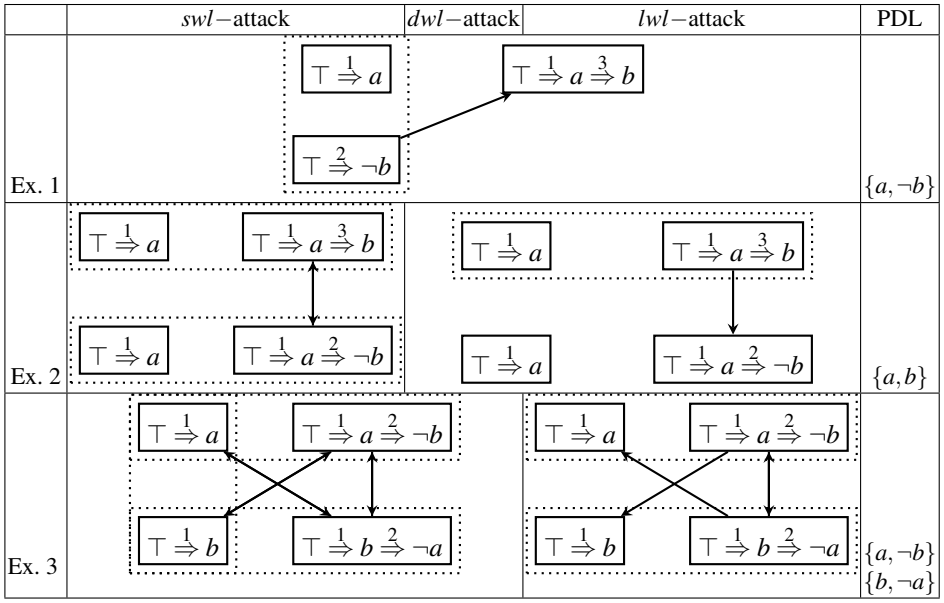
We use the formal framework for attack relation assignments introduced by Dung and Thang [12–14, 16, 17]. Their principle-based analyses of last link pointed out how weakest link must differ from last link at the level of axioms. In this paper, we propose a new lookahead weakest link attack and compare it with existing definitions also using a principle-based analysis. An important result of our paper is that the lookahead definition does not satisfy the principle of context independence [13]. We therefore introduce a new principle called weak context independence, and show that it does satisfy weak context independence. Another key result is an impossibility theorem for Dung's axioms [13] in the context of prioritised default logic [5].

Structure of the paper. Section 2 informally presents three key historical examples illustrating how to reason on weakest link. Section 3 gives the preliminary formal settings and our new attack relation. Section 4 offers the principle-based analyses. Section 5 shows that no attack relation assignment that captures PDL [5] can satisfy context independence. Section 6 discusses related work and we conclude with Section 7.

2 Three benchmark examples on weakest link

The history of weakest link evolves around three key examples which are visualised in Figure 1 and described as Examples 1–3. Note that the examples illustrate the role of formal argumentation in the context of PDL. All formal definitions are introduced later in Section 3. Here, we discuss Examples 1–3 informally.

Fig. 1: Approximating PDL in structured argumentation: a comparison of three attacks (columns) for three examples (rows). Columns are not marked when adjacent notions of attack agree on the induced attack relation at a given row. Dotted rectangles are argument extensions. Rightmost attacks approximate PDL better.



Given a knowledge base with prioritised defaults $a \overset{n}{\Rightarrow} b$ and facts (including $\top$). A prioritised default $a \overset{n}{\Rightarrow} b$ reads as: *if a then normally b*. A higher number n means a higher priority for the default rule $a \Rightarrow b$. These numerical priorities correspond to a preference relation among defaults defined by a modular order. A prioritised logic selects sets of defaults and extracts their conclusions into the so-called extensions of the logic —see Figure 2(1). A PDL extension, for example, obtains from selecting a consistent set of strongest applicable defaults. But what does a stronger *priority* mean for a default? Under the prescriptive reading, it means priority in the order of application: PDL iteratively adds the strongest applicable consistent default (Definition 18). Under the descriptive reading, the priority of a default is its contribution to the overall status of any extension containing this default [10]. The two readings clash in the most discussed example in defeasible reasoning with prioritised rules.

Example 1 (Weakest vs *last link).* Consider the three defaults: $\top \overset{1}{\Rightarrow} a$, $a \overset{3}{\Rightarrow} b$, $\top \overset{2}{\Rightarrow} \neg b$.

(Prescriptive.) One must select $\{\top \overset{2}{\Rightarrow} \neg b, \top \overset{1}{\Rightarrow} a\}$ based on application order, as shown in Figure 1. (The first choice for $\top \Rightarrow \neg b$ precludes $a \Rightarrow b$ from being selected.) This results in the extension $\{a, \neg b\}$, which is also a PDL extension.

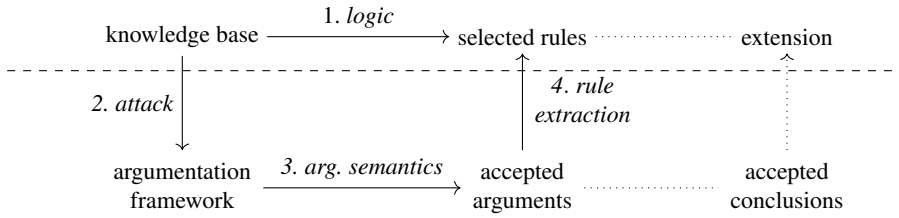


Fig. 2: Two approaches to non-monotonic inference: (1) logic systems; (2)–(4) argumentation systems. With appropriate choices on the elements (2)–(3) one can obtain exactly the same conclusions as a given logic (1).

(Descriptive.) This reading favours the set $\{\top \overset{1}{\Rightarrow} a, a \overset{3}{\Rightarrow} b\}$ as its priorities are globally better, i.e. $\{1,3\}$ vs. $\{1,2\}$. This gives the extension $\{a,b\}$, not shown in Fig. 1.[5]

Argumentation serves as a tool for representing these two interpretations of prioritised default logic using an indirect path to conclusions, as shown in Figure 2(2–4). Argumentation systems add the structure that turns collections of rules into arguments [15, 24]. An attack relation (2) among arguments, together with a semantics (3), determines the acceptance status of arguments (and their conclusions). To capture a logic, the sets of accepted arguments must correspond to the sets of defaults selected by this logic (4). **Attack relations** have thus become a major subject of study in logic-based argumentation. The direction of an attack between two conflicting arguments is often determined by their relative strengths.

Example 1 (cont'd). Suppose we want the arrow in Figure 1 (top) giving the extension $\{a,\neg b\}$ corresponding to the prescriptive reading of Example 1. This attack relation is induced by *simple weakest link* (*swl*): the strength of an argument is the lesser priority of its defaults. Under the attack relation induced by *swl*, called att_{swl}, $\top \Rightarrow \neg b$ attacks $\top \Rightarrow a \Rightarrow b$ since $2 > 1$. In fact, the three notions of weakest link considered in Figure 1 agree upon this attack relation for Example 1.

For the descriptive reading, the extension $\{a,b\}$ obtains if the attack relation is induced by *last link*, i.e. if the strength of an argument is the priority of its last default. Under last link, $\top \Rightarrow a \Rightarrow b$ attacks $\top \Rightarrow \neg b$ since $3 > 2$.[6]

For Examples 2–3, the three variants of weakest link *swl*, *dwl* and *lwl* no longer agree on the attacks or argument extensions. For each variant, Figure 1 depicts its attacks and extensions in the argumentation framework that falls under its column.

[5] Example 1, without priorities, represents the well-known Tweety scenario: *penguin* $\Rightarrow$ *bird*, *bird* $\Rightarrow$ *flies*, *penguin* $\Rightarrow \neg$*flies*. One can adduce reasons of specificity (of *penguin* over *bird*) for the standard solution: *birds fly* is overruled by the more specific rule *penguins do not fly*.

[6] These priorities give the same outputs for the *fitness-loving Scot* and *snoring professor* [23, 24], which are just variants of Example 1 with facts. Other variants of Example 1 with facts and strict rules [6, 7] give the (non-)teaching dean professor scenario [13], see Example 5 below. For further variants of Example 1 defined by partial orders we refer to Dung's paper in 2018 [16]. A brief discussion for the case of partial orders can be found in Section 7.

Example 2 (Simple vs. Disjoint weakest link). Let $\top \overset{1}{\Rightarrow} a$, $a \overset{3}{\Rightarrow} b$, $a \overset{2}{\Rightarrow} \neg b$ define our knowledge base. Note that the two arguments $\top \Rightarrow a \Rightarrow b$ and $\top \Rightarrow a \Rightarrow \neg b$ share a default with minimum priority $\top \Rightarrow a$. See the mid row in Figure 1.

(Simple weakest link.) Pollock's definition assigns the same strength 1 to these two arguments. This strength gives the mutual *swl*-attack in Figure 1 (mid, left).

(Disjoint weakest link.) A more intuitive attack relation ignores all defaults shared by two arguments in order to exploit a potential asymmetry in the remaining defaults' strengths. A relational measure of strength for such an attack is disjoint weakest link *dwl* [34]. *dwl* assigns strengths $3 > 2$ to these arguments, and generates the *dwl*-attack in Figure 1 (mid, right) that breaks the symmetry of *swl*-attacks.

Pollock's definition of weakest link *swl* [31] was adopted and studied for ASPIC+ by Modgil and Prakken [23, 24]. Young et al. [34, 35] introduced *dwl* and proved that argument extensions under the *dwl*-attack relation correspond to PDL extensions under total orders; see also the results by Liao et al. [22] or Pardo and Straßer [26]. For knowledge bases with modular orders, a new attack relation is needed for more intuitive outputs and also for a better approximation of PDL —that is, better than *dwl*.

Example 3 (Beyond dwl). Let $\top \overset{1}{\Rightarrow} a$, $\top \overset{1}{\Rightarrow} b$, $a \overset{2}{\Rightarrow} \neg b$, and $b \overset{2}{\Rightarrow} \neg a$ be the defaults.

(*swl*, *dwl*) The induced attacks admit $\{\top \Rightarrow a, \top \Rightarrow b\}$ as one of the argument extensions in Figure 1 (bottom, left). This fits neither the prescriptive interpretation nor PDL: as these two defaults are the weakest, selecting either of them ought to be followed by a stronger default, namely $a \Rightarrow \neg b$ and resp. $b \Rightarrow \neg a$. In other words, *swl* and *dwl* can select applicable defaults concurrently, leading to sub-optimal outputs.

(*lwl*) A sequential selection of defaults, more in line with PDL, is enforced by the attack relation in Figure 1 (bottom, right), induced by *lookahead weakest link* (*lwl*).

The new attack we propose (*lwl*) decides an attack from an argument by looking ahead to any superargument and its attacks: if both coincide at attacking a third argument, the former attack is disabled and only that of the superargument remains. For Example 3, this is how in Figure 1(bottom) *lwl* prevents the undesired *swl*- and *dwl*-based extension $\{a, b\}$.

3 Attack assignments based on weakest link

Preliminaries. This paper uses basic setting similar to that of Dung [13]. We assume a non-empty set $\mathcal{L}$ of ground atoms and their classical negations. An atom is also called a positive literal while a negative literal is the negation of a positive literal. A set of literals is said to be **contradictory** if it contains a pair $a, \neg a$, i.e. an atom a and its negation $\neg a$.

Definition 1 (Rule). *A defeasible rule is of the form $b_1, \ldots, b_n \Rightarrow h$ where $b_1, \ldots, b_n, h$ are domain literals. A strict rule is of the form $b_1, \ldots, b_n \rightarrow h$ where h is now either a domain literal or a non-domain atom ab_d for some defeasible rule d.*

We also define the body and head of rule r as $bd(r) = \{b_1, \ldots, b_n\}$ and $hd(r) = h$.

Instead of just assuming transitivity for the preference order among defeasible rules, as in Dung's work, in this paper we use modular orders $\preceq$ and their equivalent ranking functions *rank*. In fact, we will use the two notions indistinctly throughout the paper.

Definition 2 (Rule-based system). *A rule-based system is defined as a triple* $RBS = (RS, RD, rank)$*, where RS is a set of strict rules, RD is a finite set of defeasible rules, and rank is a function* $RD \to \mathbb{N}$ *that assigns a priority* $n = rank(d)$ *to each rule* $d \in RD$*.*

A ranking $rank : RD \to \mathbb{N}$ corresponds to a modular preorder $\preceq \subseteq RD \times RD$, i.e. a reflexive, transitive relation satisfying: $rank(d) \leq rank(d')$ iff $d \preceq d'$. A **base of evidence** BE is a (consistent) set of ground domain literals containing $\top$ and representing unchallenged facts.

Remark 1. Given the scope of our discussion and examples, our framework is less expressive than that of Dung [13]. We assume an empty set $RS = \emptyset$ of strict rules, and keep the set *RS* in Definition 2 only for notational coherence with the literature.[7]

Definition 3 (Knowledge base). *A knowledge base is a pair* $K = (RBS, BE)$ *containing a rule-based system* $RBS = (RS, RD, rank)$ *and a base of evidence* $BE \subseteq \mathcal{L} \cup \{\neg a : a \in \mathcal{L}\}$*. For convenience, we often write* $K = (RS, RD, rank, BE)$ *instead of* $K = (RBS, BE)$*.*

Example 4. The knowledge base $K = (RS, RD, rank, BE)$ for Example 3 is defined by: $RS = \emptyset$; $RD = \{d_1 : \top \Rightarrow a, d_2 : \top \Rightarrow b, d_3 : a \Rightarrow \neg b, d_4 : b \Rightarrow \neg a\}$, the function *rank* mapping $\{d_1, d_2\} \mapsto 1$ and $\{d_3, d_4\} \mapsto 2$, and finally $BE = \{\top\}$. Equivalently, we can write $K = (RS, RD, \preceq, BE)$ with $\preceq = \{d_1, d_2\}^2 \cup \{d_3, d_4\}^2 \cup (\{d_1, d_2\} \times \{d_3, d_4\})$.

Example 5 (Dean scenario). For an example with strict rules, the dean scenario asks whether the dean teaches. The knowledge base $K = (RS, RD, rank, BE)$ is given by:

$$
\begin{aligned}
RS &= \{dean \to administrator\} \\
RD &= \{dean \overset{1}{\Rightarrow} professor,\ professor \overset{3}{\Rightarrow} teach,\ administrator \overset{2}{\Rightarrow} \neg teach\} \\
BE &= \{dean\}.
\end{aligned}
$$

Definition 4 (Argument). *Given a knowledge base* $K = (RS, RD, rank, BE)$*, an* ***argument*** *wrt K is defined inductively as follows:*

1. *For each* $\alpha \in BE$*,* $[\alpha]$ *is an argument with conclusion* α*.*
2. *Let r be a rule of the form* $\alpha_1, \ldots, \alpha_n \to / \Rightarrow \alpha$ *(with* $n \geq 0$*) from K. Further suppose that* $A_1, \ldots, A_n$ *are arguments with conclusions* $\alpha_1, \ldots, \alpha_n$ *respectively. Then* $A = [A_1, \ldots, A_n \to / \Rightarrow \alpha]$*, also denoted* $A = [A_1, \ldots, A_n, r]$*, is an argument with conclusion* $cnl(A) = \alpha$ *and last rule* $last(A) = r$*.*
3. *Each argument wrt K is obtained by finitely many applications of the steps 1–2.*

[7] As a consequence, the atoms in $\mathcal{L}$ here only consist of *domain atoms* representing propositions about the concerned domains. Dung also considers *non-domain atoms* ab_d for the non-applicability of a defeasible rule d, and undercuts as strict rules $b_1, \ldots, b_n \to ab_d$ that act against the applicability of a defeasible rule d in *RD* [13]. We leave for future work the extension of our current results to knowledge bases with strict rules and undercutting arguments.

Example 6. The arguments wrt the knowledge base K from Example 4 are $A_0 = [\top]$ plus:

$$A_1 = [[\top] \Rightarrow a] \quad A_2 = [[\top] \Rightarrow b] \quad A_3 = [[[\top] \Rightarrow a] \Rightarrow \neg b] \quad A_4 = [[[\top] \Rightarrow b] \Rightarrow \neg a].$$

Definition 5 (Argumentation framework). *The set of all arguments induced by a knowledge base K is denoted by AR_K. An **argumentation framework** (AF) induced by K is a pair $AF = (AR_K, att(K))$ where $att(K) \subseteq AR_K \times AR_K$ is called an attack relation.*

Definition 6. *A knowledge base K is **consistent** if the closure of BE under RS is not a contradictory set. The set of **conclusions** of arguments in $\mathcal{E} \subseteq AR_K$ is denoted by $cnl(\mathcal{E})$.*

*A **strict** argument is an argument containing no defeasible rule. An argument is **defeasible** iff it is not strict. **The set of defeasible rules** appearing in an argument A is denoted by $dr(A)$.*

*An argument B is a **subargument** of an argument A, denoted as $B \in sub(A)$ or $B \sqsubseteq A$, iff $B = A$ or $A = [A_1, \ldots, A_n, r]$ and B is a subargument of some A_i. B is a **superargument** of A, denoted as $B \in super(A)$ or $B \sqsupseteq A$, iff $A \in sub(B)$.*

Definition 7 (Sensible class). *A class $\mathcal{K}$ of knowledge bases is **sensible** iff $\mathcal{K}$ is a non-empty class of consistent knowledge bases K, and for any knowledge base $K = (RBS, BE)$ in $\mathcal{K}$, all consistent knowledge bases of the form (RBS, BE') also belong to $\mathcal{K}$.*

Definition 8 (Attack relation assignment). *Given a sensible class of knowledge bases $\mathcal{K}$, an **attack relation assignment** is a function att mapping each $K \in \mathcal{K}$ to an attack relation $att(K) \subseteq AR_K \times AR_K$.*

Definition 9 (Stable semantics). *Given an argumentation framework $(AR_K, att(K))$, we say that $\mathcal{E} \subseteq AR_K$ is a **stable extension** if: (1) $\mathcal{E}$ is conflict-free $att(K) \cap (\mathcal{E} \times \mathcal{E}) = \emptyset$, and (2) $\mathcal{E}$ attacks all the arguments in $AR_K \setminus \mathcal{E}$. This is also denoted $\mathcal{E} \in stb(AR_K, att(K))$.*

While many other semantics exist, we follow Dung [13] and study attack relations mostly under the stable semantics. Only Principle 5 mentions the complete semantics. Recall that a set $\mathcal{E} \subseteq AR_K$ **defends** an argument A iff $\mathcal{E}$ attacks all attackers of A. A **complete** extension $\mathcal{E}$ is defined by: $\mathcal{E}$ is conflict free (no attack occurs within $\mathcal{E}$) and $A \in \mathcal{E}$ iff $\mathcal{E}$ defends A. Our main result does not depend on the choice for the stable semantics: for Examples 1–3 and the proof of Theorem 2, one can indistinctly use the complete semantics or the preferred semantics (i.e. $\subseteq$-maximally complete extensions).

Definition 10 (Belief set). *A set $S \subseteq \mathcal{L}$ is said to be a stable belief set of knowledge base K wrt an attack relation assignment att iff $att(K)$ is defined and there is a stable extension $\mathcal{E}$ of $(AR_K, att(K))$ such that $S = cnl(\mathcal{E})$.*

Attacks based on weakest link. We now present three attack relation assignments based on weakest link. All our attacks are rebuts, i.e. they contradict (sub-)conclusions. (Recall that we have neither non-domain literals nor defeasible premises that would define undercutting and resp. undermining attacks.)

Definition 11 (Contradicting attack). *Let $A, B \in AR_K$ for a knowledge base K. A **contradicts** B (at B') iff $B' \in sub(B)$ and the conclusions of A and B' are contradictory.*

Definition 12 (Weakest link). *The **weakest link** of a set of rules R, denoted as $wl(R)$, is the rank of the lowest rank rule in R. Formally, $wl(R) = \min_{r \in R} rank(r)$. Abusively, we also use $wl(A)$ for arguments A, simply defined by $wl(dr(A))$.*

Weakest link thus provides an absolute measure wl of strength for arguments—for strict arguments A, we just define $wl(A) = \infty$. This measure defines the first attack, based on Pollock's traditional idea [31].

Definition 13 (Simple weakest link attack). *Let $A, B \in AR_K$ for a knowledge base K. We say that A **swl-attacks** B (at B'), denoted as $(A,B) \in att_{swl}(K)$ iff A contradicts B at B' and $wl(A) \not< wl(B')$ (that is, $wl(A) \geq wl(B')$ for modular orders).*

Note that a defeasible argument A can contradict a strict argument B —a fact, in the present context. In those cases, $wl(A) < wl(B)$ and so the ordering $<$ is well-defined.

The second attack was introduced by Young et al. [34] for total orders. dwl was motivated by the unintuitive outputs of swl in scenarios with shared rules, like Example 2.

Definition 14 (Disjoint weakest link attack). *Let $A, B \in AR_K$ for some K. A **dwl-attacks** B (at B'), denoted $(A,B) \in att_{dwl}(K)$ iff A contradicts B at B' and $wl(dr(A) \setminus dr(B')) \not< wl(dr(B') \setminus dr(A))$.*

The third attack, newly introduced in this paper, is a refinement of disjoint weakest link. It aims to better approximate the extensions of PDL, a paradigmatic implementation of the idea of weakest link. The motivation for a new attack was given in Example 3. We call it *lookahead attack* since an attack from an argument may be cancelled if a superargument of it also attacks the same target, so this new attack looks ahead to superarguments before deciding whether an attack from the subargument ultimately exists or not.

Definition 15 (Lookahead weakest link attack). *Let $A, B \in AR_K$ for a knowledge base K. We say that $(A,B) \in att_{dwl}(K)$ is **maximal** if A is $\sqsubseteq$-maximal in AR_K with the property $(\cdot, B) \in att_{dwl}$. We also define: A **lwl-attacks** B at B', denoted as $(A,B) \in att_{lwl}(K)$, iff A dwl-attacks B at B' and*

1. *either $(B',A) \notin att_{dwl}(K)$*
2. *or, in case $(B',A) \in att_{dwl}(K)$, if (A,B) is not maximal then neither is (B',A).*

Informally, att_{lwl} obtains from att_{dwl} by removing, in each bidirectional attack, the attacker that is not $\sqsubseteq$-maximal, in case the other attacker is. With more detail, one must (1) compute $att_{dwl}(K)$; (2) for each $(A,B'), (B',A) \in att_{dwl}(K)$, if (A,B') is not maximal while (B',A) is, then remove as attacks all pairs (A,B) with $B \sqsupseteq B'$.[8]

[8] A reader might wonder why Definition 15 does not simply state: $(A,B) \in att_{lwl}(K)$ iff $(A,B) \in att_{dwl}(K)$ and A is $\sqsubseteq$-maximal with $(\cdot,B) \in att_{dwl}(K)$. The reason is that, under these attacks, one can define some K whose stable belief sets include logically contradictory sets.

Let us stress that our definition of lookahead attack *lwl* overrides the notion of contradicting attack (Definition 11). As a result, the principle of subargument structure will fail for att_{lwl}, while in general it holds for all ASPIC+ attacks in the literature.

Each of the above definitions (Defs. 13–15) of an attack relation $att(K)$ over a knowledge base K extends into an attack relation assignment *att* over a sensible class $\mathcal{K}$ of knowledge bases. This is simply the function $att : K \mapsto att(K)$ for each $K \in \mathcal{K}$.

4 Principle-based analysis

In this section, we offer a principle-based analysis of the three attack relation assignments, using the eight principles proposed by Dung [13] plus a new principle. In the following, $\mathcal{K}$ denotes a sensible class of knowledge bases, and *att* an attack relation assignment defined for $\mathcal{K}$. Some of the following results for Principles P1–P9 were partly proved by Dung [14]. With detail, our results on *swl* are also proved in Theorem 7.10 (for P1), Lemma 7.6 (for P2, P6–P8) and Theorem 7.8 (for P4).

Credulous cumulativity states that turning accepted conclusions Ω of a knowledge base K into facts preserves stable extensions and consistency. This operation is denoted as an expansion of K into $K + \Omega = (RBS, BE \cup \Omega)$.

Principle 1 (Credulous cumulativity)**.** *We say that att satisfies* ***credulous cumulativity*** *for $\mathcal{K}$ iff for each $K \in \mathcal{K}$ and each stable belief set S of K, any finite subset $\Omega \subseteq S$ satisfies:*

1. *$K + \Omega$ is a consistent knowledge base (i.e. $K + \Omega$ belongs to $\mathcal{K}$), and*
2. *S is a stable belief set of $K + \Omega$ wrt att.*

Proposition 1. *Credulous cumulativity (P1) is not satisfied by any of $att_{swl}, att_{dwl}, att_{lwl}$.*

Proof. For a counterexample, let a sensible class $\mathcal{K}$ contain the knowledge base K corresponding to Example 1. As depicted in Fig. 1(top), $S = \{a, \neg b\}$ is a stable belief set of K wrt att_{swl}, att_{dwl} and att_{lwl}. However, S is not a stable belief set of $K + \{a\}$ wrt any of these three attacks.

Context independence states that the attack relation between two arguments depends only on the rules that appear in them and their preferences [13].

Principle 2 (Context independence)**.** *We say that att satisfies* ***context independence*** *for $\mathcal{K}$ iff for any two $K, K' \in \mathcal{K}$ with preference relations $\preceq$ and resp. $\preceq'$ and any two arguments A, B belonging to $AR_K \cap AR_{K'}$, if the restrictions of $\preceq$ and $\preceq'$ on $dr(A) \cup dr(B)$ coincide, then it holds that $(A,B) \in att(K)$ iff $(A,B) \in att(K')$.*

Proposition 2. *Context independence (P2) is satisfied by att_{swl} and att_{dwl}, while it is not satisfied by att_{lwl}.*

Proof. **For att_{swl}.** Let $K, K' \in \mathcal{K}$ have preference relations $\preceq$ and resp. $\preceq'$. Suppose that for $A, B \in AR_K \cap AR_{K'}$, the restrictions of $\preceq$ and $\preceq'$ on $dr(A) \cup dr(B)$ coincide. If $(A,B) \in att_{swl}(K)$, by Def. 13 the conclusions of A and a subargument $B' \in AR_K$ of B are contradictory and $wl(A) \not\prec wl(B')$ for K. Since B' is a subargument of $B \in AR_{K'}$,

$B' \in AR_{K'}$ and $dr(B) \supseteq dr(B')$. Hence the restrictions of $\preceq$ and $\preceq'$ on $dr(A) \cup dr(B')$ also coincide. So for K' it also holds that $wl(A) \not\preceq wl(B')$. Hence, $(A,B) \in att_{swl}(K')$. The same reasoning applies in the other direction, and so we conclude that $(A,B) \in att_{swl}(K)$ iff $(A,B) \in att_{swl}(K')$.

For att_{dwl}. The proof is analogous to the proof for att_{swl}: Let $K, K' \in \mathcal{K}$ have preference relations $\preceq$ and resp. $\preceq'$. Suppose that for $A,B \in AR_K \cap AR_{K'}$, the restrictions of $\preceq$ and $\preceq'$ on $dr(A) \cup dr(B)$ coincide. If $(A,B) \in att_{dwl}(K)$, by Definition 14 the conclusions of A and a subargument $B' \in AR_K$ of B are contradictory and $wl(dr(A) \setminus dr(B')) \not\preceq wl(dr(B') \setminus dr(A))$ for K. Since B' is a subargument of $B \in AR_{K'}$, $B' \in AR_{K'}$ and $dr(B) \supseteq dr(B')$. Hence, the restrictions of $\preceq$ and $\preceq'$ on $dr(A) \cup dr(B')$ also coincide. So for K' it also holds that $wl(dr(A) \setminus dr(B')) \not\preceq wl(dr(B') \setminus dr(A))$. Hence, $(A,B) \in att_{dwl}(K')$. The same reasoning applies in the other direction, and so it holds that $(A,B) \in att_{dwl}(K)$ iff $(A,B) \in att_{dwl}(K')$.

For att_{lwl}. Let $K' = \{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b, b \overset{2}{\Rightarrow} \neg a\}$ obtain from removing $a \overset{2}{\Rightarrow} \neg b$ from the knowledge base K in Example 3. This is a counterexample, since the arguments $[\top \Rightarrow a]$ and $[\top \Rightarrow b \Rightarrow \neg a]$ belong to $AR_K \cap AR_{K'}$, and the restrictions of $\preceq$ and $\preceq'$ to the set $dr([\top \Rightarrow a]) \cup dr([\top \Rightarrow b \Rightarrow \neg a])$ coincide. However, $([\top \Rightarrow a], [\top \Rightarrow b \Rightarrow \neg a]) \notin att_{lwl}(K)$ while $([\top \Rightarrow a], [\top \Rightarrow b \Rightarrow \neg a]) \in att_{lwl}(K')$.

For a weaker version of context independence, one can state that an attack also depends on the superarguments. Let us define: $super_K(A) = \{A^+ \in AR_K : A^+ \sqsupseteq A\}$.

Principle 3 (Weak context independence)**.** *We say that att satisfies **weak context independence** for $\mathcal{K}$ iff for any two $K, K' \in \mathcal{K}$ with preferences $\preceq$ and resp. $\preceq'$ and any two arguments $A,B \in AR_K \cap AR_{K'}$:*

$$\textit{if} \left\{ \begin{array}{l} \preceq, \preceq' \ \textit{agree upon } dr(A) \cup dr(B) \\ \textit{and } \ super_K(A) = super_{K'}(A) \\ \textit{and } \ super_K(B) = super_{K'}(B) \end{array} \right\} \textit{then } (A,B) \in att(K) \textit{ iff } (A,B) \in att(K').$$

Proposition 3. *Weak context independence (P3) is satisfied by the three attacks att_{swl}, att_{dwl}, att_{lwl}.*

Proof. **For att_{swl}, att_{dwl}.** Clearly, the set of pairs $\{K, K'\}$ in $\mathcal{K}$ that need to be tested for (P3) are a subset of those pairs that to be tested for (P2): the former are all pairs validating Def. 3(i)–(ii) while the latter also include the pairs that only validate (i). Hence, if *att* satisfies (P2), then it also satisfies (P3). From this and the above proofs for (P2), we conclude that att_{swl}, att_{dwl} satisfy (P3).

For att_{lwl}. Let $K, K' \in \mathcal{K}$ have preference relations $\preceq$ and resp. $\preceq'$. Suppose that for $A,B \in AR_K \cap AR_{K'}$, $\preceq$ and $\preceq'$ agree upon $dr(A) \cup dr(B)$ and $super_K(A) = super_{K'}(A)$ and $super_K(B) = super_{K'}(B)$. Towards a contradiction, assume that $(A,B) \in att_{lwl}(K)$ at B', but $(A,B) \notin att_{lwl}(K')$. Because $(A,B) \in att_{lwl}(K)$ at B', according to Def. 15, $(A,B) \in att_{dwl}(K)$ at B'. Since att_{dwl} satisfies context independence, $(A,B) \in att_{dwl}(K')$ at B'. As a result, $(\star)$ $(B',A) \in att_{dwl}(K')$, and so (A,B) is not maximal in $att_{dwl}(K')$ and (B',A) is maximal in $att_{dwl}(K')$. Because $super_K(A) = super_{K'}(A)$ and $super_K(B) = super_{K'}(B)$, by $(\star)$ and (P3) we obtain $(B',A) \in att_{dwl}(K)$, and so (A,B) is not maximal in $att_{dwl}(K)$ and (B',A) is maximal in $att_{dwl}(K)$. Hence, $(A,B) \notin att_{lwl}(K)$. This is in contradiction with $(A,B) \in att_{lwl}(K)$.

The principle of attack monotonicity (defined below) reflects the intuition that the more reliable the foundation of an argument is, the stronger the argument becomes. Suppose the defeasible information on which an argument is based is confirmed by unchallenged observations. Replacing the defeasible bits by the observed facts should result in a strengthened argument: whatever is attacked by the original argument should also be attacked by the strengthened one, and whatever attacks the strengthened one, attacks the original one.

Definition 16 (Strengthening operation). *Let $A \in AR_K$ and $\Omega \subseteq BE$ be a finite set of domain literals. The strengthening of A wrt Ω denoted by $A \uparrow \Omega$ is defined inductively as follows:*

$$A \uparrow \Omega = \begin{cases} \{[\alpha]\} & \text{if } A = [\alpha] \text{ and } \alpha \in BE \\ AS \cup \{[hd(r)]\} & \text{if } A = [A_1, \ldots, A_n, r] \text{ and } hd(r) \in \Omega \\ AS & \text{if } A = [A_1, \ldots, A_n, r] \text{ and } hd(r) \notin \Omega \end{cases}$$

where $AS = \{[X_1, \ldots, X_n, r] \mid \forall i : X_i \in A_i \uparrow \Omega\}$

Principle 4 (Attack monotonicity)**.** *Let att be an attack relation assignment defined for a sensible class $\mathcal{K}$ of knowledge bases. We say att satisfies the property of attack monotonicity for $\mathcal{K}$ iff for each knowledge base $K \in \mathcal{K}$ and each finite subset $\Omega \subseteq BE$, the following assertions hold for arbitrary $A, B \in AR_K$ and $X \in A \uparrow \Omega$.*

1. *If $(A,B) \in att(K)$ then $(X,B) \in att(K)$.*
2. *If $(B,X) \in att(K)$ then $(B,A) \in att(K)$.*

Proposition 4. *Attack monotonicity is satisfied by att_{swl} and att_{dwl}. It is not satisfied by att_{lwl}.*

Proof. **For att_{swl}.** (1) Let $K \in \mathcal{K}$, $\Omega \subseteq BE$, $A, B \in AR_K$ and $X \in A \uparrow \Omega$. From $(A,B) \in att_{swl}(K)$, A contradicts B at some B' with $wl(A) \not< wl(B')$. Because $X \in A \uparrow \Omega$, X also contradicts B at B' with $dr(X) \subseteq dr(A)$, so $wl(X) \geq wl(A)$. As a result, $wl(X) \not< wl(B')$ and so $(X,B) \in att_{swl}(K)$. (2) From $(B,X) \in att_{swl}(K)$, B contradicts X at some X' with $wl(B) \not< wl(X')$. Because $X \in A \uparrow \Omega$, there is $A' \in sub(A)$ with $cnl(X') = cnl(A')$ and $dr(X') \subseteq dr(A')$, so $wl(X') \geq wl(A')$. As a result, B contradicts A at A' with $cnl(B)$ and $cnl(A')$ being contradictory and $wl(B) \not< wl(A')$. Thus, $(B,A) \in att_{swl}(K)$.

For att_{dwl}. The proofs are analogous to the att_{swl} case. (1) From $(A,B) \in att_{dwl}$ to $(X,B) \in att_{dwl}$: since $dr(X) \subseteq dr(A)$ we get $wl(dr(X) \setminus dr(B')) \geq wl(dr(A) \setminus dr(B')) \geq wl(dr(B') \setminus dr(A)) \geq wl(dr(B') \setminus dr(X))$. As a result, $wl(dr(X) \setminus dr(B')) \not< wl(dr(B') \setminus dr(X))$, and so $(X,B) \in att_{dwl}(K)$. (2) From $(B,X) \in att_{swl}(K)$, B contradicts X at some X' with $wl(dr(B) \setminus dr(X')) \not< wl(dr(X') \setminus dr(B))$. Because $X \in A \uparrow \Omega$, there is $A' \in sub(A)$ with $cnl(X') = cnl(A')$ and $dr(X') \subseteq dr(A')$, so $wl(dr(B) \setminus dr(A')) \geq wl(dr(B) \setminus dr(X')) \geq wl(dr(X') \setminus dr(B)) \geq wl(dr(A') \setminus dr(B))$. As a result, $(B,A) \in att_{dwl}(K)$.

For att_{lwl}. Let K contain $BE = \{a\}$ and a set RD rules of strength 1 that give: $A = [[\top \Rightarrow a] \Rightarrow b]$, $B = [[\top \Rightarrow \neg c] \Rightarrow \neg b]$, $X = [[a] \Rightarrow b]$ and also $A^+ = [A \Rightarrow c]$, $B^+ = [B \Rightarrow \neg a]$, $X^+ = [X \Rightarrow c]$. Since B^+ cannot attack X, $(A,B) \in att_{lwl}(K)$ is not preserved into $(X,B) \in att_{lwl}(K)$ although X is a strengthening of A with $\{a\}$.

The next principle, irrelevance of redundant defaults, states that adding redundant defaults into the knowledge base does not result in a change of beliefs (outputs).

Notation 1. *For any defeasible rule d, denote $K + d = (RS, RD \cup \{d\}, \preceq, BE)$ where $K = (RS, RD, \preceq, BE)$. For convenience, for any evidence $\omega \in BE$ we also denote the default $\Rightarrow \omega$ by d_ω.*

Principle 5 (Irrelevance of redundant defaults)**.** *Let $\mathcal{K}$ be a sensible class of knowledge bases such that for each $K = (RSB, BE) \in \mathcal{K}$, for each evidence $\omega \in BE$, $K + d_\omega$ belongs to $\mathcal{K}$. Further let att be an attack relation assignment defined for $\mathcal{K}$.*

We say the attack relation assignment att satisfies irrelevance of redundant defaults for $\mathcal{K}$ iff for each knowledge base $K = (RSB, BE) \in \mathcal{K}$, for each evidence $\omega \in BE$:

1. *the stable belief sets of K and $K + d_\omega$ coincide, and*
2. *the complete belief sets of K, $K + d_\omega$ coincide.*

Proposition 5. *Irrelevance of redundant defaults (P5) is satisfied by the three attacks $att_{swl}, att_{dwl}, att_{lwl}$.*

Proof. **For att_{swl}.** First, $AR_K \subset AR_{K+d_\omega}$. Let $AR^+ = AR_{K+d_\omega} \setminus AR_K$, representing arguments that are newly added into AR_{K+d_ω} due to the addition of d_ω. For each argument $A' \in AR^+$, there exists an argument $A \in AR_K$, such that $A = A' \uparrow \{\omega\}$. Hence, $cnl(A) = cnl(A')$ and $wl(A) \not< wl(A')$. Then, for each $B \in AR_K$ such that $(B, A) \in att_{swl}(K)$, we have $(B, A), (B, A') \in att_{swl}(K + d_\omega)$. Hence, A' can not be in any stable or complete extension $\mathcal{E}$ unless $A \in \mathcal{E}$. As a result, each stable or complete extension $\mathcal{E}'$ of $K + d_\omega$ is of the form $\mathcal{E} \cup \{A' \in AR^+ : A \in \mathcal{E}\}$ where $\mathcal{E}$ is an extension of K.

For att_{dwl}. The proof is analogous and only changes in statements of the form $wl(A \setminus B) \not< wl(A' \setminus B)$. Again, $cnl(A) = cnl(A')$ for any argument $A \in \mathcal{E}$ in an extension and its weakening $A' \in AR_{K+d_\omega}$. By the definition of stable and complete extensions, in the new AF these must be of the form $\mathcal{E}' = \mathcal{E} \cup \{A' : A \in \mathcal{E}\}$.

For att_{lwl}. The proof is also analogous. For each argument $A' \in AR^+$, there exists an argument $A \in AR_K$, such that $A = A' \uparrow \{\omega\}$. By definition of att_{lwl}, $(A, B) \in att_{lwl}(K)$ iff $(A, B), (A, B') \in att_{lwl}(K + d_\omega)$. As a result, each stable or complete extension $\mathcal{E}'$ of $K + d_\omega$ is of the form $\mathcal{E} \cup \{A' : A \in \mathcal{E}\}$ where $\mathcal{E}$ is an extension of K.

The next two principles state basic properties of argumentation. Subargument structure and attack closure are two basic principles. Subargument structure states that if an argument attacks a subargument, it attacks the entire argument. Attack closure says that attacks are either based on undercuts[9] or contradicting arguments.

Principle 6 (Subargument structure)**.** *Let $\mathcal{K}$ be a sensible class of knowledge bases and att be an attack relation assignment defined for $\mathcal{K}$. Then att is said to satisfy the property of subargument structure for $\mathcal{K}$ iff for each $K \in \mathcal{K}$, for all $A, B \in AR_K$,*

$(A, B) \in att(K)$ iff there is a defeasible subargument B' of B such that $(A, B') \in att(K)$.

[9] The notion of undercut from Principle 7 is the same as in Pollock [27] and ASPIC+ [24]: an argument A undercuts B at $B' \in sub(B)$ iff the last rule $d = last(B') \in RD$ and A states that this defeasible rule d is not applicable $cnl(A) = ab_d$.

Proposition 6. *Subargument structure (P6) is satisfied by att_{swl} and att_{dwl}, while it is not satified by att_{lwl}.*

Proof. **For att_{swl}.** ($\Rightarrow$) From $(A,B) \in att_{swl}(\mathrm{K})$, A contradicts some $B' \sqsubseteq B$ with $wl(A) \not< wl(B')$. If B' was strict, so would be A, contradicting that K is consistent, i.e. that $K \in \mathcal{K}$. ($\Leftarrow$) If A contradicts a defeasible B' at B'' with $wl(A) \not< wl(B'')$, then for any $B \sqsupseteq B'$ we have $wl(B) \leq wl(B') \leq wl(B'')$ and so $(A,B) \in att_{swl}(K)$.

For att_{dwl}. The two directions of the proof are analogous, now using $wl(A \setminus B') \not< wl(B' \setminus A)$ for ($\Rightarrow$); and $wl(B \setminus A) \leq wl(B' \setminus A) \leq wl(B'' \setminus A)$ for ($\Leftarrow$).

For att_{lwl}. For a counterexample to ($\Leftarrow$), let $\top \overset{2}{\Rightarrow} a$, $\top \overset{2}{\Rightarrow} \neg a$, $a \overset{1}{\Rightarrow} b$, $\neg a \overset{1}{\Rightarrow} \neg b$ be the rules of AR_K. Then, $[\top \Rightarrow a]$ does not lwl-attack $[\top \Rightarrow \neg a \Rightarrow \neg b]$ but lwl-attacks $[\top \Rightarrow \neg a]$; finally, note that $[\top \Rightarrow \neg a]$ is a subargument of $[\top \Rightarrow \neg a \Rightarrow \neg b]$.

Principle 7 (Attack closure)**.** *Let $\mathcal{K}$ be a sensible class of knowledge bases and att be an attack relation assignment defined for $\mathcal{K}$. Then att is said to satisfy the property of attack closure for $\mathcal{K}$ iff for each $K \in \mathcal{K}$, for all $A,B \in AR_K$, it holds that:*

1. *If A attacks B wrt $att(K)$ then A undercuts B or A contradicts B.*
2. *If A undercuts B then A attacks B wrt $att(K)$.*

Proposition 7. *Attack closure (P7) is satisfied by att_{swl}, att_{dwl} and att_{lwl}.*

Proof. Since we do not consider strict rules (undercuts), this principle reduces to: $(A,B) \in att$ implies A contradicts B which is immediate from Definitions 13–15.

The principle of effective rebuts enforces a natural interpretation of priorities under conflict: when two defeasible rules lead to a contradiction and so cannot be applied together, then the preferred one should be applied.

Principle 8 (Effective rebut)**.** *Let $\mathcal{K}$ be a sensible class of knowledge bases and att be an attack relation assignment defined for $\mathcal{K}$. Then att is said to satisfy the property of effective rebut for $\mathcal{K}$ iff for each $K \in \mathcal{K}$, for all $A_0,A_1 \in AR_K$ containing each exactly one defeasible rule $dr(A_0) = \{d_0\}$ and $dr(A_1) = \{d_1\}$, if A_0 contradicts A_1 then*

$$(A_0,A_1) \in att(K) \;\; \textit{iff} \;\; d_0 \not\prec d_1.$$

Proposition 8. *Effective rebut is satisfied by att_{swl} and att_{dwl}, but not by att_{lwl}.*

Proof. **For att_{swl}.** Let $dr(A) = \{d_1\}$ and $dr(B) = \{d_2\}$ contain each one defeasible rule with A,B contradicting each other. Note that $wl(A) = rank(d_1)$ and $wl(B) = rank(d_2)$. For ($\Rightarrow$), suppose that $(A,B) \in att_{swl}(K)$. As a result, A contradicts B at B' and $wl(A) \not< wl(B')$. Since $RS = \emptyset$, $B = B'$. So, $wl(A) \not< wl(B)$. That is to say, $d_1 \not\prec d_2$. For ($\Leftarrow$), suppose $d_1 \not\prec d_2$. So, $wl(A) \not< wl(B)$. Because A contradicts B at B, $(A,B) \in att_{swl}(K)$.

For att_{dwl}. The proof is analogous. Since $RS = \emptyset$, A contradicts B at B' implies $B = B'$ and $d_1 \neq d_2$. Hence, $wl(dr(A) \setminus dr(B)) = wl(A)$ and $wl(dr(B) \setminus dr(A))) = wl(B)$.

For att_{lwl}. Let AR_K contain $A = [\top \Rightarrow a]$, $B = [\top \Rightarrow \neg a]$ and $A^+ = [A \Rightarrow a]$, where all RD rules have strength 1. Then, $dr(A) = \{d_1\}$ and $dr(B) = \{d_2\}$ satisfy $d_1 \not\prec d_2$ but $(A,B) \notin att_{lwl}(K)$, since (A,B) is not maximal while (B,A) is maximal.

Attack Relation Assignment	1	2	3	4	5	6	7	8	9
swl-attack (Def. 13)	□	■	■	■	■	■	■	■	□
dwl-attack (Def. 14)	□	■	■	■	■	■	■	■	□
lwl-attack (Def. 15)	□	□	■	□	■	□	■	□	□

Table 1: Principles satisfied (■) by each attack relation assignment. Each number n refers to the Principle Pn listed next: (P1) credulous cumulativity, (P2) context independence, (P3) weak context independence, (P4) attack monotonicity, (P5) irrelevance of redundant defaults, (P6) sub-argument structure, (P7) attack closure, (P8) effective rebut, and (P9) link orientation.

The last principle, called link orientation (see below for its definition), directs attacks against those links in an argument that are identified as responsible for the argument's weakness.

Definition 17 (Weakening operation). *Let $A \in AR_K$ and $AS \subseteq AR_K$. The weakening of A by AS, denoted $A \downarrow AS$ is the set inductively defined by:*

$$A \downarrow AS = \begin{cases} \{[\alpha]\} \cup \{X \in AS : cnl(X) = \alpha\} & \text{if } A = [\alpha] \text{ and } \alpha \in BE \\ \{[X_1, \ldots, X_n, r] \mid X_i \in A_i \downarrow AS\} & \text{if } A = [A_1, \ldots, A_n, r]. \end{cases}$$

Principle 9 (Link orientation)**.** *Let $\mathcal{K}$ be a sensible class of knowledge bases and att be an attack relation assignment defined for $\mathcal{K}$. att satisfies link-orientation iff for each $K \in \mathcal{K}$, if $A, B, C \in AR_K$ are such that $C \in B \downarrow AS$, then*

$$\left\{ \begin{array}{c} (A,C) \in att(K) \text{ and} \\ \forall X \in AS, (A,X) \notin att(K) \end{array} \right\} \quad \text{implies} \quad (A,B) \in att(K).$$

That is, wrt $att(K)$, if A attacks C (the weakening of B by AS) but none of AS, then A attacks the original argument B.

Proposition 9. *Link orientation is not satisfied by any of the attacks att_{swl}, att_{dwl}, att_{lwl}.*

Proof. A counterexample for $att_{swl}, att_{dwl}, att_{lwl}$ can be found by expanding Example 1 with a new fact: $BE = \{a\}$. **For att_{swl}.** Let K consist of:

$$RD = \{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\} \quad \text{and} \quad BE = \{a\}.$$

Let $AS = \{D = [\top \Rightarrow a]\}$, $A = [\top \Rightarrow \neg b]$, $B = [[a] \Rightarrow b]$ and $C = [D \Rightarrow b]$. Note that $C \in B \downarrow AS$, and that $wl(A) = 2$, $wl(B) = 3$ and $wl(C) = 1$. Finally, observe that $(A,C) \in att_{swl}(K)$ and $(A,D) \notin att_{swl}(K)$ for $AS = \{D\}$ while $(A,B) \notin att_{swl}(K)$. **For att_{dwl}.** The same example holds, since for all the previous pairs (X,Y), $wl(X \setminus Y) = wl(X)$. **For att_{lwl}.** The same example works as in att_{dwl}, since all of A, B, C are $\sqsubseteq$-maximal attackers in K and so $att_{lwl}(K) = att_{dwl}(K)$.

Theorem 1. *The principles satisfied by each attack relation are listed in Table 1.*

Proof. This result follows from Propositions 1–9.

Discussion of the principle-based analysis. Weakest link presumes that the evaluation of an argument depends on that of its subarguments, namely their weakest components. Towards a characterization of PDL, the att_{lwl} attack relation assignment captures this idea by making the attacks from subarguments to depend on its superarguments. This results in a less compositional and more holistic view of attacks, which affects some of the principles proposed by Dung [14]. This should not be surprising at all, and instead it should be seen as part of the ongoing debate on how intuitive some of these principles are. For the popular notion of weakest link, we have a clash of intuitions. On the one hand, our intuitions on the legitimacy of weakest link and on some of our examples and, on the other, the *prima facie* intuitive principles from Dung. Following Nelson Goodman [18]'s notion of *reflective equilibrium*, this principle-based analysis should prompt us to search for a balance between intuitions on principles and intuitions on cases. Let us take a detailed look at look-ahead weakest link in Table 1.

(P1) Credulous cumulativity has also been challenged by Prakken and Vreeswijk [33, Sec. 4.4], and by Prakken and Modgil [25, Sec. 5.2]. Intuitively, the strengthened defeasible conclusion may gain the ability to defeat other arguments that they did not defeat before, which causes the stable extensions to change, thus leading to the violation of credulous cumulativity.

(P2)–(P3) Given our aim to characterize PDL and vindicate its role in non-monotonic reasoning, Context independence (P2) has to be relativized to take part of the context into account, namely the superarguments of an argument. Attack relations based on lookahead weakest link are still independent from external arguments.

(P4) The violation of one of the two directions of Attack monotonicity might be seen as the least palatable consequence of lookahead weakest link. Still, our conjecture is that the other direction (P4, item 2) holds for att_{lwl}.

(P5), (P7) The principle of Irrelevance of redundant defaults (P5) results in an intuitive property of ASPIC+, i.e. a semantic invariance under the weakening of facts into (irrelevant) defaults. Attack closure (P7) captures our understanding of how attacks in ASPIC+ should be defined. Both principles are preserved by att_{lwl}.

(P6), (P8) Despite their intuitive character, Subargument structure and Effective rebuts seem to exclude a relational notion of attacks based on the global structure around an argument, that is, the superarguments this argument is part of. The violation of these two principles might be a necessary step in any characterization of PDL in terms of attack relation assignments.

(P9) Link orientation is, in view of the counterexample in Proposition 9, one of the most disputable principle in the list. It clashes, as (P1) does, with all attack relation assignments inspired by the idea of weakest link. For anyone considering the possibility of argumentation based on weakest link, this counterexample shows that (P9) makes little sense as a general principle.

In sum, the principles proposed by Dung were inspired by last link, if not motivated towards its defense. Rather than foreclosing the existing debates on this question, we would like our principle-based analysis to open up the corresponding challenge for weakest link and its relatives, namely the search for principles that characterize the weakest link family.

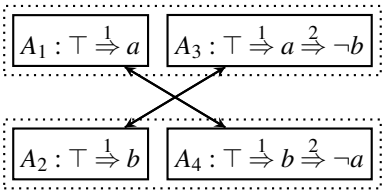


Fig. 3: AF constructed from Example 7. Arrows describe all the possible individual attacks at the subarguments. An attack relation $att(K)$ cannot contain both (A_1, A_4) and (A_2, A_3) if it is to capture the PDL extensions.

5 PDL and Dung's principles: an impossibility result

As discussed in the previous section, most of the principles proposed by Dung [14] seem indisputable, yet some others hide a partisan view on what argumentation can or cannot be. Context independence, for example, could be used to rule out Brewka's PDL from argumentation altogether. For another example, credulous cumulativity is used by Dung [13, Ex. 7.1] against elitist orderings. In turn, this principle has been further discussed and disputed by Modgil and Prakken [25].

In this section, we offer more evidence against Context independence, in the form of an impossibility result (Theorem 2). Any attempt to realize PDL in ASPIC+ should preserve the definitional principle of Attack closure (P7). Theorem 2 explains how this is incompatible with the principle of Context independence (P2).

Recall that PDL inductively applies a default of maximal priority amongst those rules that: (i) have not been applied yet, (ii) can be applied and (iii) their application does not raise an inconsistency [22, 34]. We adapt the definitions to structured argumentation.

Definition 18 (PDL). *Let $K = (RS, RD, \preceq, BE)$ be a knowledge base. For a set of defeasible rules $R \subseteq RD$, let $K{\restriction}R = (RS, R, \preceq, BE)$ and define the following sets:*

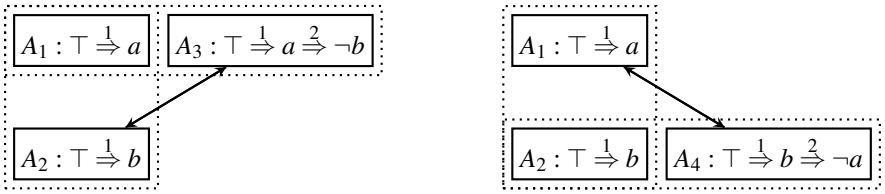
$$cl(K,R) = cnl(AR_{K{\restriction}R})$$
$$appl(K,R) = \{d \in RD \setminus R : bd(d) \subseteq cl(K,R) \text{ and } cl(K, R \cup \{d\}) \text{ is consistent}\}.$$

A PDL construction for K is any set $\bigcup_{i=0}^{\omega} R(i)$ built inductively as follows:

$$R(0) = \emptyset \qquad \text{and} \qquad R(i+1) = R(i) \cup \{d\} \quad \text{for some } d \in \max_{\preceq} appl(K, R(i))$$

*where $\max_{\preceq} \Gamma = \{d \in \Gamma \mid \forall d' \in \Gamma (d \not\prec d')\}$. Then, S is a **PDL extension** of K, denoted as $S \in pdl(K)$, if $S = cnl(K,R)$ for some PDL construction R for K.*

Example 7. Recall the set $RD = \{d_1 : \top \overset{1}{\Rightarrow} a, d_2 : \top \overset{1}{\Rightarrow} b, d_3 : a \overset{2}{\Rightarrow} \neg b, d_4 : b \overset{2}{\Rightarrow} \neg a\}$ in knowledge base $K = (RS, RD, \preceq, BE)$ from Examples 3–4. The PDL constructions for K are: $R_1 = \{d_1, d_3\}$ and $R_2 = \{d_2, d_4\}$. These constructions give the PDL extensions $S_1 = \{a, \neg b\}$ and $S_2 = \{b, \neg a\}$ respectively. Figure 3 shows an argumentation framework for K. (Note that we omit $\top$ from the PDL extensions and the argument A_0 from AR_K.)



(a) $att_1(K_1) = \{(A_2,A_3),(A_3,A_2)\}$ (b) $att_2(K_2) = \{(A_1,A_4),(A_4,A_1)\}$

Fig. 4: (a) Under the attack relation att_1, the stable extensions of $AF_1 = (AR_{K_1}, att_1)$ match the PDL extensions of K_1. (b) Similarly, for $AF_2 = (AR_{K_2}, att_2)$, we have $stb(AF_2) = pdl(K_2)$.

Example 8. Let $K_1 = (RS, RD_1, \preceq_1, BE)$ be the fragment of K consisting of $RD_1 = RD \setminus \{d_4\}$ with the preference $\preceq_1$ given by restricting $\preceq$ to the set RD_1. PDL (Def. 18) gives: $R_1 = \{d_1, d_3\} \longmapsto S_1 = \{a, \neg b\}$, and $R_3 = \{d_2, d_1\} \longmapsto S_3 = \{a, b\}$.

The PDL extensions S_1, S_3 are also obtained as (sets of the conclusions of) the stable extensions under the attack relation $att_1 = \{(A_2,A_3),(A_3,A_2)\}$. See Figure 4(a).

Example 9. Let $K_2 = \{RS, RD_2, \preceq_2, BE\}$ now be the fragment of K defined by $RD_2 = RD \setminus \{d_3\}$ and the preference $\preceq_2$ obtained by restricting $\preceq$ to RD_2. Now PDL gives: $R_2 = \{d_2, d_4\} \longmapsto S_2 = \{b, \neg a\}$, and $R_3 = \{d_1, d_2\} \longmapsto S_3 = \{a, b\}$.

The PDL extensions S_2, S_3 are also obtained as (sets of the conclusions of) the stable extensions under the attack relation $att_2 = \{(A_1,A_4),(A_4,A_1)\}$. See Figure 4(b).

Now we are in a position to prove an impossibility result for Dung's axioms and PDL, under the assumption that the axioms hold for any sensible class of knowledge bases —akin to the universal domain axiom in Arrow's impossibility theorem [1].

Theorem 2. *Let att be an attack relation assignment capturing the PDL extensions (say, under stable semantics) and satisfying attack closure (P7). Then att does not satisfy context independence (P2).*

Proof. Let $\mathcal{K}$ be a sensible class of knowledge bases containing K, K_1 and K_2 from Examples 7–9. Let also *att* be the attack relation assignment capturing the PDL extensions under stable semantics. Given this attack relation assignment *att*, the stable extensions must be the following. For $AF_0 = (AR_K, att(K))$: $\mathcal{E}_1 = \{A_1, A_3\}$ and $\mathcal{E}_2 = \{A_2, A_4\}$; for $AF_1 = (AR_{K_1}, att(K_1))$, $\mathcal{E}_1$ and $\mathcal{E}_3 = \{A_1, A_2\}$; for $AF_2 = (AR_{K_2}, att(K_2))$, $\mathcal{E}_2$ and $\mathcal{E}_3$.

The proof is by contradiction. Assume context independence (Def. 2). Using attack closure (P7), it is only the case that $\mathcal{E}_3 \in stb(AF_1)$ if $(A_2,A_3) \in att(K_1)$. Similarly, $\mathcal{E}_3 \in stb(AF_2)$ can only hold if $(A_1,A_4) \in att(K_2)$. Observe that AR_K contains all these arguments: $\{A_1,A_2,A_3,A_4\}$, and that the preference $\preceq$ from K coincides with $\preceq_1$ from K_1 on the set $\{A_1,A_2,A_3\}$ and also with $\preceq_2$ from K_2 on the set $\{A_1,A_2,A_4\}$. Hence, by context independence, we conclude that $(A_2,A_3),(A_1,A_4) \in att(K)$. But this is impossible: then $\mathcal{E}_3 = \{A_1,A_2\}$ would then become a stable extension of $AF_0 = (AR_K, att(K))$ without being a PDL extension of K. Hence, context independence is not satisfied.

6 Related work

There is a lot of work in the nonmonotonic logic and logic programming literature on prioritised rules, see e.g. Delgrande et al. [9] for an overview. Pardo and Straßer give an overview of argumentative representations of prioritized default logic, concerning weakest link, they mainly consider *dwl* [26]. Various authors discussed the dilemma between weakest link and last link [8, 22–24]. The analysis of weakest link related to *swl* indicates that it is more complicated and ambiguous than it seems at first sight. With partial orders, ASPIC+ tries to accommodate both in combination with democratic and elitist orders [23, 24], but neither of them is clearly better than the other. Young et al. [34, 35] show that even for total and modular orders, *swl* cannot always give intuitive conclusions. They also show the correspondence between the inferences made in prioritised default logic (PDL) and *dwl* with strict total orders. Then they raise the question of the similarity between weakest link and PDL for modular and partial orders. Moreover, Liao et al. [22] give similar results but use other examples to demonstrate that the approach of Young et al. [34, 35] cannot be extended to preorders [22]. Liao et al. [22] use an order puzzle in the form of Example 3 to show that even with modular orders, selecting the correct reasoning procedure is challenging. This leads them to introduce auxiliary arguments and defeats on weakest arguments. Beirlaen et al. [2] point out that weakest link is defined purely in terms of the strength of the defeasible rules used in argument construction. More recently, Lehtonen et al. present novel complexity results for ASPIC+ with preferences that are based on weakest link (*swl* in this paper) [21], they rephrase stable semantics in terms of subsets of defeasible elements.

7 Summary and future work

In this paper, we introduced a new weakest link attack relation assignment (*lwl*) and compared it with the traditional (*swl*) and disjoint (*dwl*) versions. We showed that *lwl* gets the right result for an important example (Ex. 3), at the price of loosing context independence —but this seems necessary for weakest link anyway, as shown in Table 1. As an alternative, we proposed a weaker context independence principle that is satisfied by *lwl*. A fine-grained characterization of a class of weakest link attack relation assignments, in the style of the characterizations proposed by Dung [13, 16] for last link, would also help us deepen our understanding of weakest link and vindicate its use in argumentation and non-monotonic reasoning. The core idea behind weakest link is, in our opinion, at least as important as last link for general applications in AI. On this last question, these principle-based analyses might shed some light on long-time debates between weakest link and last link, namely which one suits better each area of application of non-monotonic reasoning. Our principle-based analysis has several original insights, it presents the difference of several kinds of attack relation assignment, explains the nature of weakest link principle and reveals there is still some potential for weakest link attack to improve. By the way, it also has tight relation with some conceptual and philosophical questions and discussions: We also proved the impossibility of satisfying context independence by any attack relation assignment that captures Brewka's prioritised default logic.

As for future work, following the results presented so far, an immediate goal would be to strengthen the principle-based analysis to knowledge bases containing strict rules (and undercutting attacks). Our conjecture is that the principles satisfied by each attack relation shown in Table 1 will be preserved after the addition of strict rules. One main open question for the future of ASPIC+-style structured argumentation is which way to go: introduce auxiliary arguments like Liao et al. [22], or weaken context independence as in this paper? From a representation point of view, total orders give only one extension, while under partial orders we may have multiple extensions. Thus, another major challenge is how to generalise all the recent insights in this paper and related work to partial orders as studied in ASPIC+. While the impossibility result immediately extends from modular to partial orders, the affirmative results in our principle-based analysis need not be preserved in the latter. We thus leave for future work deciding whether this is the case for the attack relation assignments we introduced: lwl.

Finally, Table 1 also shows that the current principles fail to distinguish *swl* from dwl, while in practice they behave quite differently. Hence, another goal would be to identify a principle that separates these two attack relation assignments.

Acknowledgements. The authors are thankful to the three anonymous reviewers for their helpful comments and suggestions.

References

1. Arrow, K.J.: A difficulty in the concept of social welfare. Journal of Political Economy **54**(4), 328–346 (1950)
2. Beirlaen, M., Heyninck, J., Pardo, P., Straßer, C.: Argument strength in formal argumentation. Journal of Logics and their Applications **5**(3), 629–676 (2018)
3. Besnard, P., Hunter, A.: Elements of Argumentation. MIT Press (2008)
4. Bondarenko, A., Dung, P., Kowalski, R., Toni, F.: An abstract, argumentation-theoretic approach to default reasoning. Artificial Intelligence **93**(1), 63–101 (1997)
5. Brewka, G., Eiter, T.: Preferred answer sets for extented logic programs. Artif. Intell. **109**, 297–356 (1999)
6. Brewka, G.: Reasoning about priorities in default logic. In: Hayes-Roth, B., Korf, R.E. (eds.) Proc. of the 12th National Conference on AI. vol. 2, pp. 940–945. AAAI Press / The MIT Press (1994)
7. Brewka, G., Eiter, T.: Prioritizing default logic. In: Hölldobler, S. (ed.) Intellectics and Computational Logic. Applied Logic Series, vol. 19, pp. 27–45. Kluwer (2000)
8. Caminada, M.: Rationality postulates: Applying argumentation theory for non-monotonic reasoning. FLAP **4**(8) (2017)
9. Delgrande, J.P., Schaub, T.: Expressing preferences in default logic. Artif. Intell. **123**(1-2), 41–87 (2000)
10. Delgrande, J.P., Schaub, T., Tompits, H., Wang, K.: A classification and survey of preference handling approaches in nonmonotonic reasoning. Computational Intelligence **20**(2), 308–334 (2004)
11. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. **77**(2), 321–358 (1995)
12. Dung, P.M.: An axiomatic analysis of structured argumentation for prioritized default reasoning. Frontiers in Artificial Intelligence and Applications, vol. 263, pp. 267–272. IOS Press (2014)

13. Dung, P.M.: An axiomatic analysis of structured argumentation with priorities. Artif. Intell. **231**, 107–150 (2016)
14. Dung, P.M.: A canonical semantics for structured argumentation with priorities. In: Baroni, P., Gordon, T.F., Scheffler, T., Stede, M. (eds.) Computational Models of Argument - Proceedings of COMMA. Frontiers in Artificial Intelligence and Applications, vol. 287, pp. 263–274. IOS Press (2016)
15. Dung, P.M., Kowalski, R.A., Toni, F.: Assumption-based argumentation. In: Simari, G.R., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 199–218. Springer (2009)
16. Dung, P.M., Thang, P.M.: Fundamental properties of attack relations in structured argumentation with priorities. Artificial Intelligence **255**, 1–42 (2018)
17. Dung, P.M., Thang, P.M., Son, T.C.: On structured argumentation with conditional preferences. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI. pp. 2792–2800. AAAI Press (2019)
18. Goodman, N.: Fact, Fiction, and Forecast. Harvard University Press (1955)
19. Gorogiannis, N., Hunter, A.: Instantiating abstract argumentation with classical logic arguments: Postulates and properties. Artificial Intelligence **175**(9-10), 1479–1497 (2011)
20. Governatori, G., Maher, M.J., Antoniou, G., Billington, D.: Argumentation semantics for defeasible logic. J. Log. Comput. **14**(5), 675–702 (2004)
21. Lehtonen, T., Wallner, J.P., Järvisalo, M.: Computing stable conclusions under the weakest-link principle in the aspic+ argumentation formalism. In: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning. vol. 19, pp. 215–225 (2022)
22. Liao, B., Oren, N., van der Torre, L., Villata, S.: Prioritized norms in formal argumentation. Journal of Logic and Computation **29**(2), 215–240 (2019)
23. Modgil, S., Prakken, H.: A general account of argumentation with preferences. Artif. Intell. **195**, 361–397 (2013)
24. Modgil, S., Prakken, H.: The *ASPIC*$^+$ framework for structured argumentation: a tutorial. Argument & Computation **5**(1), 31–62 (2014)
25. Modgil, S., Prakken, H.: Abstract rule-based argumentation. In: et al., B. (ed.) Handbook of Formal Argumentation. vol. 1, pp. 287–364. College Publications (2018)
26. Pardo, P., Straßer, C.: Modular orders on defaults in formal argumentation. Journal of Logic and Computation (2022)
27. Pollock, J.L.: Defeasible reasoning. Cognitive science **11**(4), 481–518 (1987)
28. Pollock, J.L.: How to reason defeasibly. Artificial Intelligence **57**(1), 1–42 (1992)
29. Pollock, J.L.: Justification and defeat. Artificial Intelligence **67**(2), 377–407 (1994)
30. Pollock, J.L.: Cognitive carpentry: A blueprint for how to build a person. Mit Press (1995)
31. Pollock, J.L.: Defeasible reasoning with variable degrees of justification. Artif. Intell. **133**(1-2), 233–282 (2001)
32. Pollock, J.L.: Defeasible reasoning and degrees of justification. Argument Comput. **1**(1), 7–22 (2010)
33. Prakken, H., Vreeswijk, G.: Logics for defeasible argumentation. Handbook of philosophical logic pp. 219–318 (2002)
34. Young, A.P., Modgil, S., Rodrigues, O.: Prioritised default logic as rational argumentation. In: Jonker, C.M., Marsella, S., Thangarajah, J., Tuyls, K. (eds.) Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. pp. 626–634. ACM (2016)
35. Young, A.P., Modgil, S., Rodrigues, O.: On the interaction between logic and preference in structured argumentation. In: Black, E., Modgil, S., Oren, N. (eds.) Theory and Applications of Formal Argumentation - 4th International Workshop. vol. 10757, pp. 35–50. Springer (2017)