Weakest link, prioritized default logic and principles in argumentation[†]

PERE PARDO, *Department of Computer Science, University of Luxembourg, L-4364 Esch-sur-Alzette, Luxembourg.*
E-mail: pere.pardo@uni.lu

LIUWEN YU, *Department of Computer Science, University of Luxembourg, L-4364 Esch-sur-Alzette, Luxembourg.*
E-mail: liuwen.yu@uni.lu

CHEN CHEN, *School of Philosophy, Zhejiang University, Hangzhou 310007, China.*
E-mail: 12104018@zju.edu.cn

LEENDERT VAN DER TORRE, *Department of Computer Science, University of Luxembourg, L-4364 Esch-sur-Alzette, Luxembourg.*
School of Philosophy, Zhejiang University, Hangzhou 310007, China.
E-mail: leon.vandertorre@uni.lu

Abstract

In this article, we study procedural and declarative logics for defaults in modular orders. Brewka's prioritized default logic (PDL) and structured argumentation based on weakest link are compared to each other in different variants. This comparison takes place within the framework of attack relation assignments and the axioms (principles) recently proposed for them by Dung. To this end, we study which principles are satisfied by weakest link and disjoint weakest link attacks. With the aim of approximating PDL using argumentation, we identify an attack defined from PDL extensions, prove that each such PDL extension is a stable belief set under it and offer a similar principle-based analysis. We also prove an impossibility theorem for Dung's axioms that covers PDL-inspired attack relation assignments. Finally, a novel variant of PDL with concurrent selection of defaults is also proposed, and compared to these argumentative approaches. In sum, our contributions fill an important gap in the literature created by Dung's recent methods and open up new research questions on these methods.

Keywords: logic, closure, knowledge base, selected, extension, framework.

[†] This article is a major expansion of our work [11] presented at CLAR 2023. The main novelties are listed next: we extend principle-based analyses to knowledge bases with strict rules, in line with Dung [16]; we revise our previous work on *lookahead weakest link* with a PDL-based attack relation assignment, prove it contains all PDL extensions and also study its principles. Finally, we introduce a concurrent version of PDL (pPDL), compare it to previous methods and also discuss its impact on key examples for defeasible reasoning.

Vol. 35, No. 4, © The Author(s) 2025. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.
https://doi.org/10.1093/logcom/exaf007

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[1] was used by John Pollock in 1995 as a way to compare the strength of arguments:

> 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. [35, p. 99]

Defeasible arguments build from prioritized rules or **defaults** $a \overset{n}{\Rightarrow} b$, reading *if a, then normally b*, where a higher number n means a **higher priority or strength**. A central question in defeasible reasoning is how an argument draws strength from its defaults: is it the strength of its *weakest link* or that of its *last link*? (*Last link* claims that the strength of an argument is that of its last default.)[2] This dilemma also affects normative scenarios, where a rule $a \Rightarrow b$ can instead represent a norm for an obligation b. The choice for *weakest link* or *last link* has thus an impact on queries to prioritized knowledge bases and normative systems: *Do fitness-loving Scots like whisky? Should snoring professors get access to the library?* [27, 28]:

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

The distinction between *weakest* and *last link* played a central role in formal models of structured argumentation, whose foundations were laid by Pollock in a series of influential articles [33–37].[3] Around the same time, Dung published a seminal paper on abstract argumentation that became as well part of the foundations of formal argumentation [14]. This general framework provides semantics of attacks for structured argumentation, as in the ASPIC+ system [27, 28], and has been used to instantiate a variety of non-monotonic logics [14, 22]. Of particular interest is the case of prioritized default logic (PDL) [7], a logic closely related to weakest link [42]. More recently, the focus has shifted from argument strength to the attack relations—and the axioms to be imposed on them [15, 16].

For an example, consider the knowledge base containing the defaults and fact:

$$\{\top \overset{1}{\Rightarrow} a, \quad \top \overset{2}{\Rightarrow} \neg b, \quad a \overset{3}{\Rightarrow} b\ \} \qquad \text{and} \qquad \top$$

(This is essentially the *fitness-lover Scot* example, with priorities.) Say a default $x \Rightarrow y$ is **applicable** if its antecedent x has been derived as (or is) a fact. Starting with the set of facts $\{\top\}$, the **closure** of this set under applicable defaults gives another set $\{\top, a, b, \neg b\}$, a contradictory output. Different methods remedy this situation by computing consistent outputs for such defeasible knowledge bases.

[1] Thomas Reid [41] wrote in 1786: *In every chain of reasoning, the evidence of the last conclusion can be no greater than that of the weakest link of the chain, whatever may be the strength of the rest.'*

[2] For a sample of views on this dilemma: Prakken and Sartor [39] claim that *last link* applies in legal reasoning, while Amgoud and Cayrol [1] choose *weakest link* for handling inconsistencies. The axiomatic approach started by Dung [16, 17] can be naturally read as a defense for *last link*.

[3] This important distinction between *weakest* and *last link* necessitates the possibility of representing default rules $a \Rightarrow b$. This distinction cannot be made in argumentation systems without defaults, such as Assumption-Based Argumentation (ABA) [5] or classical logic-based argumentation [4].

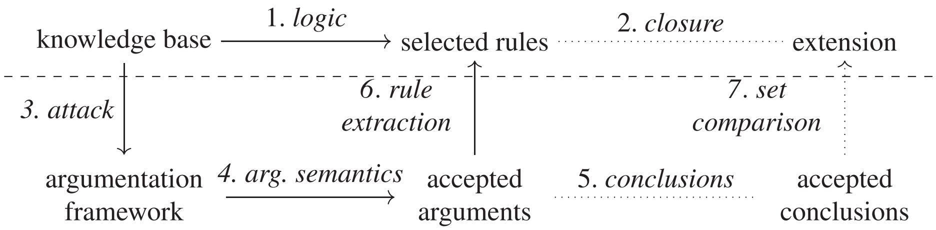


FIGURE 1. Two approaches to non-monotonic inference: (1)–(2) logics; (3)–(5) argumentation systems. With appropriate choices on items (3)–(4) (e.g. weakest link), one can obtain the same defaults (6) and/or outputs (7) as a given logic (e.g. PDL)

These methods evolved from (I) logics to (II) argumentation systems (see Figure 1 for the relation between I and II), and more recently into (III) attack relations.

(I. Logics of defeasible reasoning: PDL.) These logics work by selecting sets of defaults that are maximally consistent—see Figure 1(1). PDL, in particular, iteratively selects a strongest applicable default. In the example, this selection is

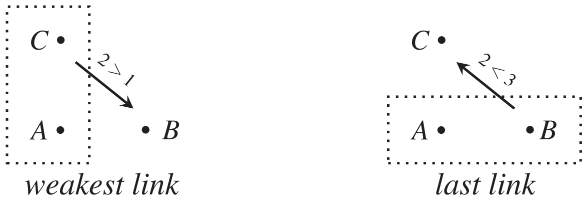
$$\top \overset{2}{\Rightarrow} \neg b \qquad \text{and then} \qquad \top \overset{1}{\Rightarrow} a$$

which gives the PDL extension $\{a, \neg b\}$—see Figure 1(2). Note that $a \overset{3}{\Rightarrow} b$ becomes applicable now, but PDL does not select it as it would give a contradiction.

(II. Argumentation: Weakest Link) With the same defeasible knowledge base, ASPIC+ builds instead an argument A, B, C for each applicable default:

$$A : \top \overset{1}{\Rightarrow} a \qquad B : \top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b \qquad C : \top \overset{2}{\Rightarrow} \neg b$$

This isolates the conflict $\{b, \neg b\}$ as taking place between the two arguments $\{B, C\}$. Conflicts like this are to be resolved by defining directed attack(s)—see Figure 1(3). *Weakest link* induces the attack $C \to B$, while *last link* gives $B \to C$, as shown next:



Once an attack relation is fixed, an argumentation semantics selects arguments that form a self-defending set[4] —see also Figure 1(4). In the graphs above, the selected arguments are $\{A, C\}$ under *weakest link*; and $\{A, B\}$ under *last link*. Observe that the defaults of $\{A, C\}$ match the PDL selection $\{\top \Rightarrow \neg b, \top \Rightarrow a\}$—Figure 1(6). And the conclusions of $\{A, C\}$ match the PDL extension $\{a, \neg b\}$—see Figure 1(5,7).

For this scenario, weakest link captures PDL, and so it exemplifies the general schema (Figure 1) describing how an argumentation system captures a logic.

[4] A set X is *self-defending* if it provides counterattacks $A' \to B \to A$ for each attack $B \to A$ against a member $A \in X$. That is, X contains an attacker A' for each attacker B against some $A \in X$

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) attack monotonicity, (P4) irrelevance of redundant defaults, (P5) subargument structure, (P6) attack closure, (P7) effective rebut, (P8) link orientation and (P9) weak context independence

Attack Rel. Assignment	1	2	3	4	5	6	7	8	9
swl-attack (Definition 13)	□	■	■	■	■	■	■	□	■
dwl-attack (Definition 14)	□	■	■	■	■	■	■	□	■
pdl-attack (Definition 16)	□	□	■	■	■	■	□	■	□
lwl-attack [11]	□	□	□	■	□	■	□	□	■

$$\begin{tikzcd}
\mathrm{PDL} \arrow[r] \arrow[dr] \arrow[ddr] & \mathrm{pPDL} \\
 & att_{pdl} \\
att_{dwl} \arrow[r] & att_{swl}
\end{tikzcd}$$

FIGURE 2. Inclusions between default logics and attack-based argumentation

(III. Axiomatic methods) Attack relations have thus become a major subject of study in logic-based argumentation. Dung [15, 16] recently proposed an axiomatic method that applies to all argumentation systems with defeasible rules.[5] The idea is that among all candidates to being an attack relation of a knowledge base, namely all binary relations, only those satisfying the axioms qualify as an attack.

Along this line, principle-based analyses [9, 15, 16, 18, 21] have studied general properties of attack relations under various approaches to structured argumentation.[6] However, given the long history of weakest link, it may come as a surprise that there have been few developments characterizing how it can be used to instantiate abstract argumentation frameworks (AFs) that capture a given logic.

Starting with traditional weakest link [27, 28, 35], and the variant called disjoint weakest link [42], we explain this saga and its relation to PDL using three benchmark examples. We then study the **research questions**: (Q1) what axioms for attack relations are satisfied by each variant of weakest link? (See Table 1 in Section 5.) (Q2) how do attacks and logics that approximate weakest link relate to each other? (See Sections 4,7, Figure 2 or Table 2.) These two questions also aim to shed light on the dilemma between weakest and last link.

Our **methodology** is the formal framework for attack relation assignments introduced by Dung and Thang [15–19]. Our **contributions** include:

- a study of axioms for weakest link attacks att_{swl}, att_{dwl} under total preorders;
- a similar study of a new attack relation att_{pdl}, based on PDL;
- a comparitive study with a new, concurrent variant of PDL, called pPDL;

[5] This claim must be qualified for the ASPIC+ system, which also defines (1) defeasible premises $\Rightarrow a$ and (2) contrary functions $\overline{a} = \{\neg a, b, \ldots\}$ that generalize classical negation. These elements, though, are often represented as (1) defaults $\top \Rightarrow a$, and resp. (2) strict rules $b \rightarrow \neg a$, etc. See [28, Section 4.6]. For this reason, we will indistinctly refer to the ASPIC+ system or Dung's framework.

[6] And earlier than this focus in attack relations, studies proposed principles for the logical behaviour of structured argumentation systems, in the form of consistency and closure postulates [10].

TABLE 2. A comparison of PDL, pPDL and three attack relations. If each output (extension of K, stable belief set under $att(K)$) of the row method is also an output of the column method, the table lists the formal result proving this. Otherwise, the Table describes an example disproving this inclusion. An interrogation mark '?' denotes an open problem

	PDL	pPDL	att_{dwl}	att_{swl}	att_{pdl}
PDL	—	Fact 7.1	Example 8	Theorem 4.3	Theorem 4.4
pPDL	Example 3	—	Example 8	?	Example 3
att_{dwl}	Example 3	?	—	Fact 4.1	Example 3
att_{swl}	Example 2	Example 2	Example 2	—	Example 2
att_{pdl}	Example 9	?	Example 8	?	—

- a partial hierarchy for these logics and attack relations (Figure 2) as a step towards bridging the gap between procedural and declarative approaches;
- an impossibility theorem for Dung's axioms [16] in the context of PDL [7].

Structure of the paper. Section 2 presents three key historical examples illustrating how to reason with weakest link. Section 3 provides the formal preliminaries. Section 4 studies attack relation assignments and their relation to PDL. Section 5 offers principle-based analyses for all these attacks. Section 6 shows the impossibility of any attack relation matching PDL [7] to satisfy the axiom of context independence. Section 7 studies a concurrent variant of PDL. Section 8 discusses related work and Section 9 concludes with a summary and future work.

2 Three benchmark examples on weakest link

The history of weakest link evolves around three key examples from the literature, visualized in Figure 3 and described as Examples 1–3. All formal definitions are introduced later in Sections 3–4. Here, we discuss Examples 1–3 informally. Each example tests which weakest link variants approximate PDL better.

The first example is about priorities: what does a *stronger priority* mean? Under the **prescriptive** reading [13], it means priority in the order of application. Under the **descriptive** reading, the priority means how desirable it is that the default is applied. Recall the *fitness-loving Scot* example. The two readings clash in the most discussed example in defeasible reasoning with prioritized rules.[7]

EXAMPLE 1 (Weakest *vs* Last link).
Let $\{\top \overset{1}{\Rightarrow} a, a \overset{3}{\Rightarrow} b, \top \overset{2}{\Rightarrow} \neg b\}$ be again our defaults (Figure 3, top). The two readings of priorities give the following outputs:

(*Prescriptive*.) Based on application order, one must select $\{\top \overset{2}{\Rightarrow} \neg b, \top \overset{1}{\Rightarrow} a\}$ thereby obtaining the output $\{a, \neg b\}$, as in PDL. In fact, PDL (Definition 15) is an implementation of the prescriptive reading. Let us call *simple weakest link* (*swl*) the strength defined by the lowest priority of an argument:

$$\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b \quad \longmapsto \quad 1 = \min\{1,3\} \qquad\qquad \top \overset{2}{\Rightarrow} \neg b \quad \longmapsto \quad 2 = \min\{2\}$$

[7] Other variants of this example with facts and strict rules [6, 8] give the *dean* scenario [16]—see Example 5 below. For further variants defined by partial orders we refer to Dung's paper in 2018 [18] or our brief discussion in Section 9.

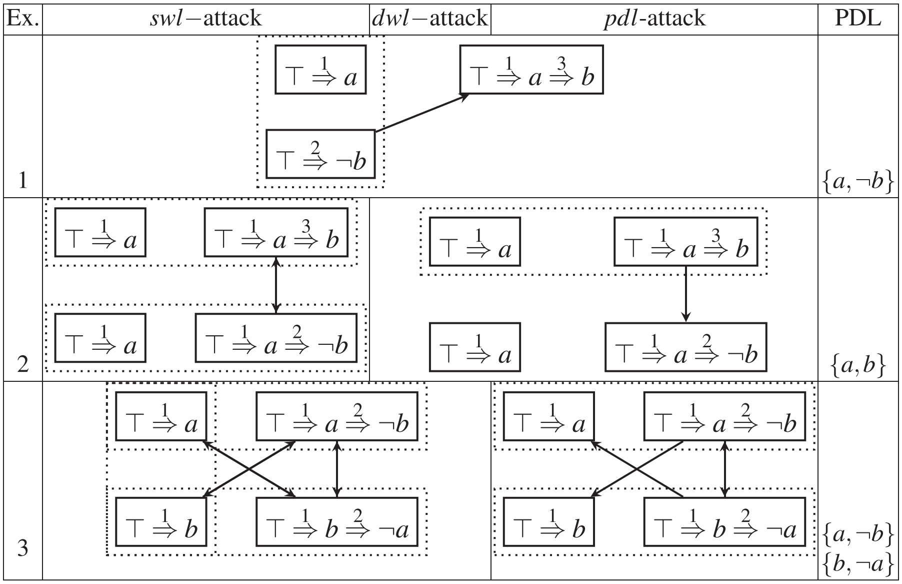

Ex.
$swl-$attack
$dwl-$attack
pdl-attack
PDL
1
$\top \overset{1}{\Rightarrow} a$
$\top \overset{2}{\Rightarrow} \neg b$
$\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$
$\{a, \neg b\}$
2
$\top \overset{1}{\Rightarrow} a$
$\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$
$\top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$
$\{a, b\}$
3
$\top \overset{1}{\Rightarrow} a$
$\top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$
$\top \overset{1}{\Rightarrow} b$
$\top \overset{1}{\Rightarrow} b \overset{2}{\Rightarrow} \neg a$
$\{a, \neg b\}$
$\{b, \neg a\}$

FIGURE 3. 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

A comparison of the strengths in this conflict produces the attack shown in Figure 3 (top). The semantics then gives the argument selection also shown. Our three attack relations (swl, dwl, pdl) do in fact agree on the verdict for this example.[8]

(*Descriptive*.) This reading favours the set $\{\top \overset{1}{\Rightarrow} a, a \overset{3}{\Rightarrow} b\}$ as its priorities $\{1,3\}$ are more desirable than the rival ones $\{1,2\}$. *Last link* can be seen as an implementation of this reading: the contribution of a new default to a selection or argument, say $\{\top \Rightarrow a\}$, is defined by the desirability of this default (2 vs. 3 in the example). Last link thus agrees on the above preference but arrives at it through argumentative means. First, one computes argument strength:

$$\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b \quad \longmapsto \quad 3 = \text{last}(1,3) \qquad\qquad \top \overset{2}{\Rightarrow} \neg b \quad \longmapsto \quad 2 = \text{last}(2)$$

Based on this, argument $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$ attacks $\top \overset{2}{\Rightarrow} \neg b$. Using a standard argumentation semantics, one obtains the output $\{a, b\}$, not shown in Figure 3 (top).

Simple weakest link does not always capture the prescriptive reading though. In response to this, a more intuitive *disjoint* variant of weakest link has been considered [42]. This relational measure of argument strength ignores all defaults shared by two arguments, and compares the weakest links of their disjoint fragments only.

[8] This example represents the Tweety scenario $\{penguin \rightarrow bird, bird \Rightarrow flies, penguin \Rightarrow \neg flies\}$ with priorities instead of the strict rule ($\rightarrow$). Without priorities, the solution $\{penguin, bird, \neg flies\}$ obtains from specificity (of *penguin* over *bird*): *birds fly* is overruled by the more specific *penguins do not fly*. Without specificity the solution obtains from appropriate priorities using PDL or *swl*.

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. The two conflicting arguments $\top \Rightarrow a \Rightarrow b$ and $\top \Rightarrow a \Rightarrow \neg b$ share a default $\top \Rightarrow a$ with the lowest priority. See the mid row in Figure 3.

(Simple weakest link.) Pollock's definition assigns the same strength 1 to these two arguments. This gives the mutual *swl*-attack in Figure 3 (mid, left). One argument selection $\top \Rightarrow a \Rightarrow b$ matches the PDL extension $\{a, b\}$; the other $\top \Rightarrow a \Rightarrow \neg b$, though, gives us a non-PDL extension $\{a, \neg b\}$.

(Disjoint weakest link.) The attack relation defined by *disjoint weakest link* (*dwl*) assigns strengths $3 > 2$ to the above arguments, after excluding the default they share. This generates the tie-breaking *dwl*-attack shown in Figure 3 (mid, right). This figure also shows the set of arguments selected by our semantics. The selected arguments' conclusions match the PDL output $\{a, b\}$.

Pollock's definition of weakest link *swl* [36] was adopted and studied for ASPIC+ by Modgil and Prakken [27, 28]. Young et al. [42, 43] 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. [24] and Pardo and Straßer [30]. Under total preorders, a new attack relation is needed for more intuitive outputs and a better approximation of PDL—that is, better than *dwl*.

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

(*swl*, *dwl*) Weakest link attacks, depicted in Figure 3 (bottom, left), admit the selection of arguments $\{\top \Rightarrow a, \top \Rightarrow b\}$. This selection neither fits the prescriptive interpretation nor PDL: selecting either default ought to be followed by the selection of a stronger default, namely $a \Rightarrow \neg b$ and resp. $b \Rightarrow \neg a$.

(PDL) As PDL selects one strongest default at a time, this excludes by construction the concurrent selection of $\{\top \Rightarrow a, \top \Rightarrow b\}$. The PDL-inspired attack relation (Definition 16) in Figure 3 (bottom, right) also excludes this selection.[9]

3 Preliminaries

We use the formal setting of Dung [16] for the present section. We assume a non-empty set $At = \{a, \ldots\}$ of ground atoms and their classical negations in a language $\mathcal{L}$, which also contains the *true* constant $\top$. An atom a is also called a positive literal while a negative literal $\neg a$ is the negation of a positive literal. A set of literals is said to be **contradictory** if it contains a pair $a, \neg a$.

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 either a domain literal or a non-domain atom ab_d for some

[9] The same attack relation for Example 3 is induced by a further variant *lookahead weakest link* (*lwl*) [11]. As pointed out by C. Straßer, this variant *lwl* does not satisfy Subargument Closure. For this reason, instead of *lwl*, we study here an attack relation built directly from PDL (Definition 16).

defeasible rule d. For any rule r, we define the body of r as $bd(r) = \{b_1, \ldots, b_n\}$ and the head of r as $hd(r) = h$.

The **language** $\mathcal{L}$ consists of the set $At \cup \{\neg a : a \in At\} \cup \{\top\} \cup \{ab_d, \ldots\}$ containing a *non-domain atom* ab_d for each defeasible rule d. The atom ab_d states the non-applicability of this rule d. A strict rule of the form $b_1, \ldots, b_n \rightarrow ab_d$ is then used to build undercuts against any argument that makes use of rule d.

Rather than assuming partial orders for preferences among defeasible rules, as in Dung's work, we focus on total preorders $\preceq$, also called modular orders.[10] The equivalent notion of ranking functions *rank* will also be used indistinctly.

DEFINITION 2 (Rule-based system).
A rule-based system $RBS = (RS, RD, \preceq)$ consists of a set RS of strict rules, a finite set RD of defeasible rules and a total preorder $\preceq$ on RD. Equivalently, we write $RBS = (RS, RD, rank)$ with a function $rank : RD \rightarrow \mathbb{N}$ satisfying $rank(d) \leq rank(d')$ iff $d \preceq d'$.

A **base of evidence** is a non-contradictory set $BE \subseteq At \cup \{\neg a : a \in At\} \cup \{\top\}$. It contains ground domain literals and $\top$, but no non-domain atoms of the form ab_d. A base of evidence BE represents unchallenged facts.

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. We will also 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

$$RS = \emptyset \qquad RD = \left\{ \begin{array}{l} d_1 : \top \Rightarrow a \\ d_2 : \top \Rightarrow b \\ d_3 : a \Rightarrow \neg b \\ d_4 : b \Rightarrow \neg a \end{array} \right\} \qquad \begin{array}{l} d_1 \xrightarrow{rank} 1 \\ d_2 \longrightarrow 1 \\ d_3 \longrightarrow 2 \\ d_4 \longrightarrow 2 \end{array} \qquad BE = \emptyset$$

Equivalently, the total preorder $\preceq = \{d_1, d_2\}^2 \cup \{d_3, d_4\}^2 \cup (\{d_1, d_2\} \times \{d_3, d_4\})$ defines the knowledge base $K = (RS, RD, \preceq, BE)$.

EXAMPLE 5 (Dean scenario).
For an example with strict rules, the dean scenario asks whether the dean teaches or not. Let $K = (RS, RD, rank, BE)$ be given by

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

[10] A modular order is a reflexive and transitive relation (a preorder) that is also total: for any elements $d, d' \in RD$ either $d \preceq d'$ or $d' \preceq d$ (or both).

This scenario is structurally similar to Example 1. Weakest link concludes $\neg teach$, while last link opts for *teach*—and both accept $\{dean, administrator, professor\}$.

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

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

EXAMPLE 6.
The arguments wrt the knowledge base K from Example 4 are

$$\begin{array}{lll} A_0 = [\top] & A_1 = [[\top] \Rightarrow a] & A_2 = [[\top] \Rightarrow b] \\ & A_3 = [[[\top] \Rightarrow a] \Rightarrow \neg b] & A_4 = [[[\top] \Rightarrow b] \Rightarrow \neg a]. \end{array}$$

DEFINITION 5 (Argumentation framework).
The set of all arguments induced by a knowledge base K is denoted by AR_K. An **AF** (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, denoted $Cl_{RS}(BE)$, is not a contradictory set. The set of **conclusions** of arguments in a subset $\mathcal{E} \subseteq AR_K$ 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 strict (resp. defeasible rules) appearing in an argument A is denoted by $sr(A)$ (resp. $dr(A)$).

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.

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 such $K = (RBS, BE)$ 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 $att : K \mapsto att(K)$ mapping each $K \in \mathcal{K}$ to an attack relation $att(K) \subseteq AR_K \times AR_K$ such that no attack $(A, B) \in att(K)$ exists against a strict argument $B \in AR_K$.

Definition 9 (Stable semantics).
Given an AF $(AR_K, att(K))$, we say that $\mathcal{E} \subseteq AR_K$ is a **stable extension**, denoted $\mathcal{E} \in stb(AR_K, att(K))$, if:

(1) $\mathcal{E}$ is conflict-free: $att(K) \cap (\mathcal{E} \times \mathcal{E}) = \emptyset$, and
(2) $\mathcal{E}$ attacks all of $AR_K \setminus \mathcal{E}$: for each $B \notin \mathcal{E}$ there is $A \in \mathcal{E}$ with $(A, B) \in att(K)$.

While many other semantics exist, we follow Dung [16] and study attack relations mostly under the stable semantics. Only principle (P4) mentions the complete semantics. A set $\mathcal{E} \subseteq AR_K$ **defends** an argument A iff $\mathcal{E}$ attacks all attackers of A. $\mathcal{E}$ is a **complete** extension if $\mathcal{E}$ is conflict-free and $A \in \mathcal{E}$ iff $\mathcal{E}$ defends A.

Our results do not depend on the choice for stable semantics: e.g. for Examples 1–3 and the proof of Theorem 6.1, one can indistinctly use the complete semantics, or the preferred semantics defined by $\subseteq$-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 $S = cnl(\mathcal{E})$.

Given an argument A, a **basic defeasible subargument** of A is any subargument $A' \in sub(A)$ whose last rule is defeasible $last(A') \in RD$.
Attack types. When defining an attack relation assignment based on (disjoint) weakest link, each attack will either be a *rebut* contradicting a conclusion, or an *undercut* whose conclusion ab_d aims at any argument using d as its last rule.[11]

Definition 11 (Undercut, rebut).
Let $A, B \in AR_K$ for a knowledge base K. We say that

A **undercuts** B at B' iff $B' \in sub(B)$, $last(B') = d$ is defeasible and $cnl(A) = ab_d$.
A **contradicts** B at B' iff $B' \in sub(B)$ and the conclusions of A and B' are contradictory: either $cnl(A) = \neg cnl(B')$ or $cnl(B') = \neg cnl(A)$.[12]
A **rebuts** B at B' iff A contradicts B at B', where B' is a basic defeasible subargument; that is, $last(B') \in RD$.

Let us observe that Dung [16] first considers assignments defined by *undercuts* and *contradicting attacks*. Under the principle of Credulous Cumulativity, it is shown that stable extensions remain the same if one considers instead *undercuts* and *rebuts* [16, Lemma 5.1], the so-called *basic attacks*.

But Credulous Cumulativity fails for all of our attack relation assignments (Proposition 5.1), so we directly adopt basic attacks for our principle-based analyses. Under appropriate assumptions (e.g. closure of *RS* under transposition), our analyses remain the same if attacks consist of undercuts and contradicting attacks.

[11] Following [16], our language does not include defeasible premises and so we need not define undermining attacks against them. (See also fn. 5).

[12] Observe that contradictory conclusions are always pairs of domain literals, that is, not of not of the form ab_d.

4 Attack relations based on weakest link and PDL

Attacks based on weakest link. We first present two attack relation assignments that measure argument strength as simple and disjoint weakest link.

DEFINITION 12 (Weakest link).
Let $R \subseteq RD$ be a set of defeasible rules. The **weakest link** of R, denoted $wl(R)$, is the rank of the lowest ranked defeasible rule in R. Formally, $wl(R) = \min_{r \in R} rank(r)$. We abusively extend $wl(\cdot)$ to arguments A:

$$wl(A) = \begin{cases} \infty & \text{if } A \text{ is strict} \\ wl(dr(A)) & \text{if } A \text{ is defeasible.} \end{cases}$$

Like last link, weakest link provides an absolute measure of argument strength. Unlike last link, wl does not depend on proof structure: it evaluates arguments as unstructured sets of defaults. We rephrase Pollock's traditional idea [36] as follows.

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 $B' \in sub(B)$, denoted $(A, B) \in att_{swl}(K)$, iff

$$A \text{ undercuts } B \text{ at } B' \quad \text{or} \quad (A \text{ rebuts } B \text{ at } B' \text{ and } wl(A) \not< wl(B')).$$

Observe that for modular orders $wl(A) \not< wl(B')$ reduces to $wl(A) \geq wl(B')$. Note also that a defeasible argument A can contradict a strict argument B, but any such case will verify $wl(A) < wl(B) = \infty$. The ordering $<$ is thus well-defined.

The second attack relation assignment, called disjoint weakest link dwl, was introduced by Young et al. [42] for total orders. dwl was motivated by the unintuitive outputs of swl in scenarios with shared rules, such as Example 2. dwl provides a (comparative) measure of an argument's strength in relation to another argument:

$$A <_{dwl} B \quad \text{if}\{f\} \quad wl\left(dr(A) \setminus dr(B)\right) < wl\left(dr(B) \setminus dr(A)\right)$$

DEFINITION 14 (Disjoint weakest link attack).
Let $A, B \in AR_K$ for some K. We say that A **dwl-attacks** B at $B' \in sub(B)$, denoted $(A, B) \in att_{dwl}(K)$, iff

$$A \text{ undercuts } B \text{ at } B' \quad \text{or} \quad (A \text{ rebuts } B \text{ at } B' \text{ and } A \not<_{dwl} B').$$

The attack relation $att(K)$ defined over each K under Definitions 13–14 extends into an attack relation assignment att defined over any sensible class $\mathcal{K} = \{K, \dots\}$ of knowledge bases. These attack relation assignments are defined by the maps:

$$att_{swl} : K \longmapsto att_{swl}(K) \quad \text{and} \quad att_{dwl} : K \longmapsto att_{dwl}(K).$$

FACT 4.1. Stable extensions under att_{dwl} are also stable extensions under att_{swl}. □

PROOF. For a given $K = (RS, RD, \preceq, BE)$, let $\mathcal{E} \in stb(AR_K, att_{dwl}(K))$.
($\mathcal{E}$ is conflict free under $att_{swl}(K)$.) Suppose the contrary towards a contradiction. An attack $(A, B) \in att_{swl}(K)$ would be preserved as an attack within $\mathcal{E}$ in either direction: $(A, B) \in att_{dwl}(K)$ or $(B, A) \in att_{dwl}(K)$, contradicting the assumption that $\mathcal{E}$ is a stable extension of K under $att_{dwl}(K)$.
($\mathcal{E}$ swl-attacks all of $AR_K \setminus \mathcal{E}$.) Let $B^+ \in AR_K \setminus \mathcal{E}$ be arbitrary, and let $A \in \mathcal{E}$ be a dwl-attacker of B^+ at some B; that is, $(A, B^+) \in att_{dwl}(K)$. We prove that $(A, B^+) \in att_{swl}(K)$ as well. If the dwl-attack is an undercut, then the same undercut exists under $att_{swl}(K)$ for the arguments. If the dwl-attack is a rebut, then we have $A \geq_{dwl} B$. Clearly, if $dr(A) \cap dr(B) = \emptyset$, then we must have $wl(A) \geq wl(B)$ and so $(A, B^+) \in att_{swl}(K)$. Otherwise, let $C_1, \ldots, C_q$ be $\sqsubseteq$-maximal subarguments of both A and B, and let k be their minimum rank $k = \min\{wl(C_1), \ldots, wl(C_q)\}$. Using the inequality $wl(dr(A) \setminus dr(B)) \geq wl(dr(B) \setminus dr(A))$, consider the cases

(Case $k \geq wl(dr(A) \setminus dr(B))$.) Then we obtain that $wl(A) = wl(dr(A) \setminus dr(B)) \geq wl(dr(B) \setminus dr(A)) = wl(B)$. From $wl(A) \geq wl(B)$, $(A, B^+) \in att_{swl}(K)$.

(Case $wl(dr(A) \setminus dr(B)) > k$.) This implies the inequalities: $wl(dr(A) \setminus dr(B)) \geq wl(A) = k \geq min\{k, wl(dr(B) \setminus dr(A))\} = wl(B)$. From these, one obtains again that $wl(A) \geq wl(B)$ and so $(A, B^+) \in att_{swl}(K)$.

In both cases, we showed that $(A, B^+) \in att_{swl}(K)$. Since $B^+ \notin \mathcal{E}$ was arbitrary, $\mathcal{E}$ swl-attacks all of $AR_K \setminus \mathcal{E}$. □

Attacks based on PDL. Let us start with a brief reminder of PDL. For the sake of simplicity, we rewrite the original terminology and notation to those used above for argumentation. The correspondence is as follows:[13]

a prioritized default theory	K	a knowledge base over At only
a normal default $\frac{a\,:\,b}{b}$	$a \Rightarrow b$	a defeasible rule in RD
claims known	$RS \cup BE$	strict rules and facts
a total preorder $\preceq$	$\preceq$ or $rank$	a total preorder or a ranking.

REMARK 1.
Prioritized default theories K are only defined over a set At of domain atoms. As non-domain atoms ab_d are not defined, strict rules in RS do not give rise to undercutting attacks. In this section, knowledge bases K will be of this form.

We denote the deductive **closure** of a set E under rules by $Cl_{RS}(E)$. A rule $d = \phi \Rightarrow \psi$ is **active** in a set E if $\phi \in E$ and $\psi, \neg\psi \notin E$.

DEFINITION 15 (PDL).
Given a prioritized default theory $K = (RS, RD, \preceq, BE)$, a **PDL extension** is a set $E = \bigcup_i E_i$ inductively defined by

$$\begin{aligned} E_1 &= Cl_{RS}(BE) \\ E_{n+1} &= \begin{cases} Cl_{RS}(E_n \cup hd(r_{n+1})) & \text{if condition } (\star) \text{ below holds} \\ E_n & \text{otherwise, if } (\star) \text{ fails for all } \phi \end{cases} \end{aligned}$$

[13] The general definition of a default is an expression of the form $\frac{a\,:\,c}{b}$, reading: *if we know a and what we know is consistent with c, then infer b*. The logic PDL [7] applies to this general concept of defaults, while in the present article we focus on normal defaults. Normal defaults, which are defaults of the form $\frac{a\,:\,b}{b}$, are here simply called defaults.

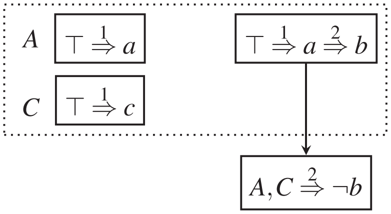

A
$\top \overset{1}{\Rightarrow} a$
$\top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} b$
C
$\top \overset{1}{\Rightarrow} c$
$A, C \overset{2}{\Rightarrow} \neg b$

FIGURE 4. For Example 8, the stable belief set under att_{dwl} is $\{a, c, b\}$. Two extensions $\{a, c, b\}$ and $\{a, c, \neg b\}$ exist under PDL (and pPDL in Definition 21)

where ($\star$) is the consequent ϕ of a $\preceq$-maximal rule $r_{n+1} \in RD$ active in E_n exists that makes $Cl_{RS}(E_n \cup \phi)$ non-contradictory.

To each PDL extension E corresponds a PDL *construction* $R = \bigcup_n R_n$ given by $R_1 = \emptyset$ and $R_{n+1} = R_n \cup \{r_{n+1}\}$, where r_{n+1} is the default (if any) from E_{n+1}.

An important research question is then to characterize, or at least to approximate, the PDL extensions of a prioritized default theory $K = (RS, RD, \preceq, BE)$ using an attack relation $att(K)$ over the arguments from K. For total orders[14] $\preceq = \leq$, an attack that characterizes PDL extensions already exists: att_{dwl}.

THEOREM 4.2 (Young et al. [42, Theorem 5.3]).
For total orders $\preceq$, E is a (unique) PDL extension of K iff $E = cnl(\mathcal{E})$ for a (unique) stable extension $\mathcal{E}$ of (AR_K, att_{dwl}).

For total preorders, though, it is an open problem how to characterize PDL extensions using an attack relation assignment. But certainly such a characterization can no longer be based on disjoint weakest link.

EXAMPLE 7 (Disjoint weakest link *vs* PDL).
Example 3 shows a stable belief set $\{a, b\}$ under $att_{dwl}(K)$ that is not a PDL extension of K.

EXAMPLE 8 (PDL *vs* Disjoint weakest link).
Define $K = (RS, RD, rank, BE)$ by

$$RS = \emptyset \qquad RD = \left\{ \begin{array}{ll} \top \overset{1}{\Rightarrow} a, & a \overset{2}{\Rightarrow} b \\ \top \overset{1}{\Rightarrow} c, & a, c \overset{2}{\Rightarrow} \neg b \end{array} \right\} \qquad BE = \emptyset$$

As shown in Figure 4, the shared rule $\top \overset{1}{\Rightarrow} a$ produces only one stable extension $\mathcal{E}$ under disjoint weakest link, and so a unique stable belief set of $(AR_K, att_{dwl}(K))$:

$$\mathcal{E} = \{A, C, [A \Rightarrow b]\} \quad \longmapsto \quad S = \{a, b, c\}$$

[14] Recall that a total order $\leq$ over RD is reflexive ($d \leq d$), transitive ($d \leq d' \leq d'' \Rightarrow d \leq d''$), antisymmetric ($d \leq d' \leq d \Rightarrow d = d'$) and strongly connected ($d \leq d'$ or $d' \leq d$). The strict part of a total order $\leq$ is the strict total order $<$ defined by its irreflexive fragment $< \, = \, \leq \setminus Id_{RD}$.

In contrast, two PDL constructions exist for K, and so do two PDL extensions:

$$\begin{array}{lll} \left(\top \overset{1}{\Rightarrow} a, \quad a \overset{2}{\Rightarrow} b, \quad \top \overset{1}{\Rightarrow} c\right) & \longmapsto & \{a,b,c\} \\ \left(\top \overset{1}{\Rightarrow} c, \quad \top \overset{1}{\Rightarrow} a, \quad a,c \overset{2}{\Rightarrow} \neg b\right) & \longmapsto & \{a,\neg b,c\} \end{array}$$

As a consequence, disjoint weakest link cannot characterize PDL under the stable semantics. Observe that att_{swl} here coincides with PDL.

As shown next, for total preorders, PDL is approximated better by weakest link than by its disjoint variant. At least, att_{swl} provides an upper bound for PDL.

THEOREM 4.3.
Let $K = (RS, RD, \preceq, BE)$ be a prioritized default theory. For any PDL extension E of K, E is a stable belief set under $att_{swl}(K)$.

PROOF. Let E be a PDL construction of K, with PDL construction R. We extend the set of rules R so as to include as well-redundant rules:

$$R^+ = R \cup \{d \in RD \setminus R : bd(d) \cup \{hd(d)\} \subseteq R\}.$$

Define $K' = (RS, R^+, \preceq', BE)$, where $\preceq'$ is the restriction of $\preceq$ to R^+, and then the set of arguments $\mathcal{E} = AR_{K'}$. Clearly, $cnl(\mathcal{E}) = E$.
($\mathcal{E}$ is conflict free.) From the fact that E is non-contradictory, and that $\mathcal{E}$ is closed under subarguments, it follows that no rebuts exist within $\mathcal{E}$. As a consequence, no attacks $(A, B) \in att_{swl}(K)$ exist for any $A, B \in \mathcal{E}$.

($\mathcal{E}$ attacks all of $AR_K \setminus \mathcal{E}$.) Let $B^+ \in AR_K \setminus \mathcal{E}$ be arbitrary. By the maximality of the PDL construction R, there must be a rule $d \in dr(B^+)$ such that $E \cup \{hd(d)\}$ is contradictory. Let $B \sqsubseteq B^+$ be an argument with such a rule $last(B) = d$, and $\sqsubseteq$-minimal with this property. That is, with $bd(d) \subseteq E$. Now suppose towards a contradiction that $\mathcal{E}$ does not swl-attack B at B. From the fact that $E \cup \{hd(d)\}$ is contradictory, there exist some arguments $A_1, \ldots, A_m$ in $\mathcal{E}$ that rebut B at B. By definition of att_{swl}, it must be that $wl(A_1), \ldots, wl(A_m) < wl(B)$. Let $d_1 \in dr(A_1), \ldots, d_m \in dr(A_m)$ be arbitrary weakest links in these arguments. Since $d_1, \ldots, d_m \prec dr(B)$, the PDL construction R for E will make each weakest link d_j active, but not select it since some other default in $dr(B)$, among others, has priority provided that $d_j \prec dr(B)$. Thus, R is not a PDL construction for K, a contradiction. □

The converse of Theorem 4.3 does not hold, in view of Example 2.

For another approximation of PDL, let us define an attack relation $att_{pdl}(K)$ directly from the PDL extensions of K. This also provides an upper bound for PDL, in the sense that all PDL extensions are stable belief sets under $att_{pdl}(K)$.

DEFINITION 16 (PDL attack).
Let $A, B \in AR_K$ for some $K = (RS, RD, rank, BE)$. We say that A **pdl-attacks** B at a subargument B', denoted $(A, B) \in att_{pdl}(K)$, iff A undercuts B or there is a PDL extension E, with PDL construction R, such that

(i) $dr(A) \subseteq R$ (ii) $bd(last(B')) \subseteq E$ and (iii) A rebuts B at B'.

THEOREM 4.4.
Let $K = (RS, RD, \preceq, BE)$ be a prioritized default theory. For any PDL extension E of K, E is a stable belief set under $att_{pdl}(K)$.

PROOF. Let E be a PDL extension of K, with PDL construction R. We extend this set R with the redundant defaults at E, and define

$$R^+ = R \cup \{d \in RD \setminus R : bd(d) \cup \{hd(d)\} \subseteq R\}.$$

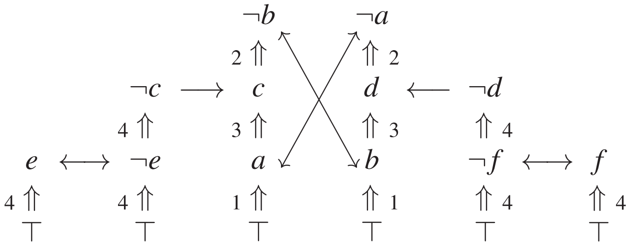
Consider then the prioritized default theory $K' = (RS, R^+, \preceq', BE)$, where $\preceq'$ is the restriction of $\preceq$ to R^+, and define the extension $\mathcal{E} = AR_{K'}$. We prove that $\mathcal{E}$ is stable.

($\mathcal{E}$ is conflict-free.) This follows from E being non-contradictory, and *pdl*-attacks, of the form (A, B) at B', being defined by contradictory pairs of conclusions $cnl(A), cnl(B')$. (Observe that $\mathcal{E} = AR_{K'}$ is closed under subarguments).

($\mathcal{E}$ attacks all of $AR_K \setminus \mathcal{E}$.) Let $B \in AR_K \setminus \mathcal{E}$. Hence, there must be $d \in dr(B)$ such that $d \notin R^+$. (Otherwise, we would have $B \in \mathcal{E}$.) Let then $B' \sqsubseteq B$ be a $\sqsubseteq$-minimal argument with this property; that is, $d = last(B') \notin R^+$. Since $R^+ \supseteq R$, we also have that $d \notin R$. By the minimality of B', for each subargument $B'' \sqsubset B'$, we must have $dr(B'') \subseteq R^+$, so that $hd[dr(B'')] \subseteq E$, and since this holds for arbitrary $B'' \sqsubset B'$, we obtain (1) $bd(d) \subseteq E$. Let n be minimum with the property that $hd[dr(B'')] \subseteq E_n$ for all $B'' \sqsubset B'$. By the maximality of PDL constructions, there must be some default d_{n+k} (at a step $R_{n+k} \subseteq R$) such that (2) $hd(d_{n+k})$ contradicts $hd(d)$. Hence, let $A \in \mathcal{E}$ be such that $last(A) = d_{n+k}$. Such an argument must exist since $d_{n+k} \in R$. Clearly $dr(A) \subseteq R^+$ but, moreover, we can choose such A to satisfy in fact (3) $dr(A) \subseteq R$. From (2), A rebuts B at B'. Combining this with (1) $bd(d) \subseteq E$ and (3) $dr(A) \subseteq R$, Definition 16 gives us that $(A, B) \in att_{pdl}(K)$. □

The converse of Theorem 4.4 does not hold, as shown next.

EXAMPLE 9 (pdl-attack *vs* PDL).
Consider the prioritized default theory K inducing the next arguments. (For $att_{pdl}(K)$, we only depict those attacks $X \longrightarrow Y$, where X pdl-attacks Y at Y. All attacks on superarguments of Y are omitted).



In the framework above, there exist 8 stable belief sets: seven of them are PDL extensions of K; but the stable belief set $\{a, b, c, d, e, f\}$ is not a PDL extension.

5 Principle-based analyses

This section offers an analysis of each attack relation assignment att_{swl}, att_{dwl} and att_{pdl} using the eight principles P1–P8 proposed by Dung [16], plus a new principle P9 that weakens P2.[15] In the

[15] For simple weakest link att_{swl}, some of the following results already follow from those proved by Dung [16]: our result for P1 is also proved as [16, Thm 7.10]; for principles P2 and P5–P7, as [16, Lemma 7.6]; and finally for P3, as [16, Thm

following, $\mathcal{K}$ denotes a sensible class of knowledge bases, and *att* an attack relation assignment defined for $\mathcal{K}$.

Credulous cumulativity (P1). This principle states that the operation of turning accepted conclusions Ω of a knowledge base K into facts preserves both its stable belief sets and its consistency. This operation is expressed as an expansion of K into $K + \Omega = (RBS, BE \cup \Omega)$.

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

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

Credulous cumulativity does not generally hold for non-monotonic rule-based systems [26], a fact that was already well known before the foundational articles in argumentation. It should not come as a surprise that this principle does fail for all three attacks here considered. See also the discussion at the end of this section.

PROPOSITION 5.1.
Credulous cumulativity (P1) is not satisfied by any of the attack relation assignments $att_{swl}, att_{dwl}, att_{pdl}$.

PROOF. **For att_{swl}, att_{dwl}.** For a counterexample, let a sensible class $\mathcal{K}$ contain the knowledge base K corresponding to Example 1. As depicted in Figure 3 (top), $S = \{a, \neg b\}$ is a stable belief set of K wrt att_{swl}, att_{dwl}. However, S is not a stable belief set of $K + \{a\}$ wrt any of att_{swl}, att_{dwl}. The stable extension $\mathcal{E}'$ contains now:

$$[a] \qquad [[\top] \overset{1}{\Rightarrow} a] \qquad [[a] \overset{3}{\Rightarrow} b]$$

which defeat the argument $[[\top] \overset{2}{\Rightarrow} \neg b]$. The stable belief set (and PDL extension) of $K + \{a\}$ is then $S' = \{a, b\}$. S' differs from the original $S = \{a, \neg b\}$.

For att_{pdl}. We reason with same knowledge bases K (left) and $K + \{a\}$ (right):

$$\begin{array}{rll}
R & = \{\top \overset{2}{\Rightarrow} \neg b, \top \overset{1}{\Rightarrow} a\} & = \{a \overset{3}{\Rightarrow} b, \top \overset{1}{\Rightarrow} a\} \\
att_{pdl}(\cdot) & = \{([[\top] \overset{2}{\Rightarrow} \neg b], [[[\top] \overset{1}{\Rightarrow} a] \overset{3}{\Rightarrow} b])\} & = \{([[a] \overset{3}{\Rightarrow} b], [[\top] \overset{2}{\Rightarrow} \neg b])\} \\
cnl(\mathcal{E}) & = \{a, \neg b\} & = \{a, b\}
\end{array}$$

Hence, the stable belief set $\{a, \neg b\}$ for K is not preserved into one for $K + \{a\}$. □

Context independence (P2). This principle states that the attack relation between two arguments depends only on the strengths of rules that appear in them.

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$(A, B) \in att(K')$. □

7.8]. We include our own proofs in order to make this contribution self-contained. The operations defined in Defs. 17–19 are also taken from [16].

PROPOSITION 5.2.
Context independence (P2) is satisfied by att_{swl} and att_{dwl}. It is not satisfied by att_{pdl}.

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. ($\subseteq$.) Let $(A,B) \in att_{swl}(K)$. By Definition 13 either (1) A undercuts B at some B', or (2) the conclusions of A and some basic defeasible subargument $B' \in AR_K$ of B are contradictory and these arguments satisfy $wl(A) \not\prec wl(B')$ for K. (1) For the undercut case, we immediately obtain that $(A,B) \in att_{swl}(K')$ also holds. (2) For the rebut case, since B' is a subargument of $B \in AR_{K'}$, we also have that B' is in $AR_{K'}$ and obviously satisfies $dr(B) \supseteq dr(B')$. Hence, the restrictions of $\preceq$ and $\preceq'$ on $dr(A) \cup dr(B')$ also coincide. So for K', it also holds that $wl(A) \not\prec wl(B')$ for some basic defeasible subargument B' of B. In both cases (1)–(2), we obtained that $(A,B) \in att_{swl}(K')$, and so we are done. ($\supseteq$.) For the other inclusion an analogous reasoning shows the claim. From this, we conclude that $(A,B) \in att_{swl}(K)$ iff$(A,B) \in att_{swl}(K')$.

For att_{dwl}**.** The proof is similar to the previous claim. 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. ($\subseteq$.) If $(A,B) \in att_{dwl}(K)$, again by Definition 14 the undercut case is immediate. For the rebut case, the conclusions of A and a basic defeasible subargument $B' \in AR_K$ of B are contradictory and $A \not\prec_{dwl} B'$ 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 $A \not\prec_{dwl} B'$ for some basic defeasible B'. Hence, $(A,B) \in att_{dwl}(K')$ holds for both the undercut and rebut cases. ($\supseteq$.) The same reasoning applies in the other direction. In conclusion, so it holds that $(A,B) \in att_{dwl}(K)$ iff$(A,B) \in att_{dwl}(K')$.

For att_{pdl}**.** Let $K = (RS, RD, \preceq, BE)$ and $K' = (RS, RD', \preceq', BE)$ be:

$$
\begin{array}{lllll}
(K) & RS = \emptyset & RD = \{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg a, \top \overset{3}{\Rightarrow} a\} & & BE = \{\top\} \\
(K') & RS = \emptyset & RD' = \{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg a\} & & BE = \{\top\}
\end{array}
$$

Let $A = [[\top] \overset{1}{\Rightarrow} a]$, $B = [[\top] \overset{2}{\Rightarrow} \neg a]$ and, in AR_K, $C = [[\top] \overset{3}{\Rightarrow} a]$. The (unique) PDL construction for K is $R = \{\top \overset{3}{\Rightarrow} a\}$ and for K' is $R' = \{\top \overset{2}{\Rightarrow} \neg a\}$. Then,

- the arguments A, B belong to $AR_K \cap AR_{K'}$,
- $\preceq$ and $\preceq'$ coincide within the set $dr(A) \cup dr(B)$, but
- $(B,A) \notin att_{pdl}(K) = \{(C,B)\}$, while $(B,A) \in att_{pdl}(K')$.

Hence, (P2) fails. □

Attack monotonicity (P3). This principle 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, in the following sense: 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 17 (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 by

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

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

PRINCIPLE 3 (Attack monotonicity). Let *att* be an attack relation assignment defined for a sensible class $\mathcal{K}$ of knowledge bases. We say that *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 5.3.
Attack monotonicity (P3) is satisfied by att_{swl}, att_{dwl} and att_{pdl}.

PROOF. **For** att_{swl}**.** (1) Let $K \in \mathcal{K}$, $\Omega \subseteq BE$, $A, B \in AR_K$ and $X \in A \uparrow \Omega$. Suppose that $(A, B) \in att_{swl}(K)$. (Case: rebut.) That is, A contradicts B at some basic defeasible subargument $B' \sqsubseteq B$ with $wl(A) \not< wl(B')$. Because $X \in A \uparrow \Omega$, X also contradicts B at B' with $dr(X) \subseteq dr(A)$, and so $wl(X) \geq wl(A)$. From this and the above inequation we obtain $wl(X) \geq wl(A) \geq wl(B')$. As a result, $wl(X) \not< wl(B')$ and so $(X, B) \in att_{swl}(K)$. (Case: undercut.) There is $B' \in sub(B)$ such that $last(B') = d \in RD$ and $cnl(A) = ab_d$. We have that $ab_d \notin BE$ as it is not a domain literal, and so any $X \in A \uparrow \Omega$ satisfies $cnl(X) = ab_d$. In conclusion, X also undercuts B at B'.

(2) Let $(B, X) \in att_{swl}(K)$ be arbitrary. (Case: rebut.) In this case B contradicts X at some basic defeasible subargument $X' \sqsubseteq X$ with $wl(B) \not< wl(X')$. Because $X \in A \uparrow \Omega$, there is a subargument $A' \sqsubseteq A$ satisfying $cnl(X') = cnl(A')$ and such that $dr(X') \subseteq dr(A')$. Using the first property, $A' \sqsubseteq A$ must also be basic defeasible, while the second implies $wl(X') \geq wl(A')$. As a result, B contradicts A at a basic subargument $A' \sqsubseteq A$ with $cnl(B)$ and $cnl(A')$ being contradictory and such that $wl(B) \not< wl(A')$. Thus, $(B, A) \in att_{swl}(K)$. (Case: undercut.) Then there is $X' \sqsubseteq X$ with $d = last(X')$ and $cnl(B) = ab_d$. By [16, Lemma 4.2], we obtain that $X' \in A' \uparrow \Omega$ for some subargument $A' \sqsubseteq A$. And clearly $X' \neq [\alpha]$ for any $\alpha \in \Omega \cup hd[dr(A)]$. Hence, by construction (Definition 17), X' is of the form $X' = [X'_1, \ldots, X'_m, d]$. That is, $d = last(X') = last(A')$ and so B also undercuts A at A' since $cnl(B) = ab_d$.

For att_{dwl}**.** (1) From $(A, B) \in att_{dwl}$ to $(X, B) \in att_{dwl}$. (Case: rebut.) Hence, X attacks B at a basic defeasible subargument $B' \sqsubseteq B$. Using the fact that $dr(X) \subseteq dr(A)$ (twice) and the case assumption we reason as follows:

$$wl(dr(X) \setminus dr(B')) \geq wl(dr(A) \setminus dr(B')) \geq wl(dr(B') \setminus dr(A)) \geq \\ \ldots \geq wl(dr(B') \setminus dr(X)).$$

As a result, $wl(dr(X) \setminus dr(B')) \not< wl(dr(B') \setminus dr(X))$, that is $X \not<_{dwl} B'$, and since $B' \sqsubseteq B$ is basic defeasible, we conclude that $(X, B) \in att_{dwl}(K)$. (Case: undercut.) The proof is exactly as in the claim for att_{swl}.

(2) Let $(B, X) \in att_{swl}(K)$. (Case: rebut.) B contradicts X at some basic defeasible subargument X' with $wl(dr(B) \setminus dr(X')) \not< wl(dr(X') \setminus dr(B))$. Because $X \in A \uparrow \Omega$, again by [16, Lemma 4.2] there is $A' \in sub(A)$ with $cnl(X') = cnl(A')$ and $dr(X') \subseteq dr(A')$. Again the former property implies

that (1) $A' \sqsubseteq A$ is basic defeasible. The latter, together with the above inequality implies (2):

$$wl(dr(B) \setminus dr(A')) \geq wl(dr(B) \setminus dr(X')) \geq wl(dr(X') \setminus dr(B)) \geq \\ \ldots \geq wl(dr(A') \setminus dr(B)).$$

Finally, (1)–(2) jointly imply that $(B, A) \in att_{dwl}(K)$. (Case: undercut.) The proof is again as in the case for att_{swl}.

For att_{pdl}**.** (1) Let $(A, B) \in att_{epdl}(K)$. Hence, there is a PDL extension E of K such that $hd[dr(A)] \subseteq E$ and a subargument $B' \in sub(B)$ such that A rebuts B at B' and (1) $bd(last(B')) \subseteq E$. Let R be the PDL construction for E. Then, by definition of X, (2) $dr(X) \subseteq dr(A) \subseteq R$. Since $cnl(X) = cnl(A)$, the set $\{cnl(X), cnl(B')\}$ is also contradictory and so (3) X rebuts B at B'. Putting (1)–(3) together gives us that $(X, B) \in att_{pdl}(K)$.

(2) Let now $(B, X) \in att_{pdl}(K)$. Thus, there is a PDL construction R such that (1) $dr(B) \subseteq R$, and B rebuts X at some X' with $bd(last(X')) \subseteq E$. Then there is a subargument $A' \sqsubseteq A$ with $cnl(A') = cnl(X')$ for which (2) B also rebuts A at A' and, moreover, (3) $bd(last(A')) = bd(last(X')) \subseteq E$. From (1)–(3), we conclude that $(B, A) \in att_{pdl}(K)$. □

Irrelevance of redundant defaults (P4). Let us call a default redundant[16] if its conclusion is already an established fact in a knowledge base. The next principle states that adding redundant defaults does not result in a change of beliefs.

DEFINITION 18.
For any defeasible rule d, let $K + d = (RS, RD \cup \{d\}, \preceq, BE)$ denote the expansion of a knowledge base $K = (RS, RD, \preceq, BE)$ with a defeasible rule d. For any evidence $\omega \in BE$, the default of the form $d_\omega = \top \Rightarrow \omega$ is called redundant.

PRINCIPLE 4 (Irrelevance of redundant defaults). Let $\mathcal{K}$ be a sensible class such that for each $K = (RBS, BE) \in \mathcal{K}$ and each fact $\omega \in BE$, $K + d_\omega$ belongs to $\mathcal{K}$. We say that att (defined over such $\mathcal{K}$) satisfies irrelevance of redundant defaults for $\mathcal{K}$ iff for each knowledge base $K = (RBS, BE) \in \mathcal{K}$ and 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 5.4.
Irrelevance of redundant defaults (P4) is satisfied by att_{swl}, att_{dwl} and att_{pdl}.

PROOF. **For** att_{swl} **and** att_{dwl}**.** Dung proved that Attack Monotonicity (P3) and Context Independence (P2) imply the principle Irrelevance of Redundant Defaults (P4) [16, Theorem 4.1]. From this and Propositions 5.2–5.3, the claim follows.

For att_{pdl}**.** Note that for any PDL extension E, $\omega \in BE \subseteq E$, and so for any PDL construction R, the redundant default d_ω is not in R. We use Definition 19 below.

(Stable semantics.) It is clear that an argument A is in a stable extension of K iff both A and $A \downarrow \{d_\omega\}$ are in a a stable extension of $K + \{d_\omega\}$. (Otherwise, $A \downarrow \{d_\omega\}$ would be attacked by this extension, making the former extension not conflict-free.) In the following, let $A' \in \{A, A \downarrow \{d_\omega\}\}$ and similarly for B'. For each PDL extension E (with construction R) of K, there is a PDL extension

[16] This meaning of 'redundant' is different from that we used in the proof of Theorem 4.4.

E' (and resp. R') of $K + d_\omega$ such that the following equivalences hold:

$$\begin{array}{rcll}
dr(A) \subseteq R & \text{if}\{f\} & dr(A) \subseteq R' & \\
bd(last(B)) \subseteq E & \text{if}\{f\} & bd(last(B')) \subseteq E' & \\
(B,A) \in att_{pdl}(K) & \text{if}\{f\} & (B,A') \in att_{pdl}(K + d_\omega) & (\textit{Definition } 16, \text{any } B \in AR_K) \\
(A,B) \in att_{pdl}(K) & \text{if}\{f\} & (A,B') \in att_{pdl}(K + d_\omega) & (\textit{Definition } 16, \text{any } B \in AR_K).
\end{array}$$

And vice versa, each PDL extension E' of $K + \{d_\omega\}$ also obtains from a PDL extension E of the form above. Thus, there is a correspondence between stable extensions $\mathcal{E}$ of K and $\mathcal{E}'$ of $K + \{d_\omega\}$, where $\mathcal{E}' = \mathcal{E} \cup \{A \downarrow \{d_\omega\} : A \in \mathcal{E}\}$. Since $cnl(A) = cnl(A \downarrow \{d_\omega\})$, the stable belief sets of K and $K + \{d_\omega\}$ coincide.

(Complete semantics.) The proof is analogous, except that now we use the completeness of $\mathcal{E}$: $A \in \mathcal{E} \Leftrightarrow \mathcal{E}$ defends $A \Leftrightarrow \mathcal{E}'$ defends $A' \Leftrightarrow A' \in \mathcal{E}'$. Again, the complete extensions of K and $K + d_\omega$ only differ on the $A \downarrow \{d_\omega\}$ arguments. Since $cnl(A) = cnl(A \downarrow \{d_\omega\})$, the complete belief sets of K and $K + \{d_\omega\}$ coincide. □

Subargument structure (P5). Subargument structure states that if an argument attacks a subargument, it also attacks the entire argument, and vice versa.

PRINCIPLE 5 (Subargument structure). An attack relation assignment *att* for $\mathcal{K}$ satisfies the property of subargument structure iff for each $K \in \mathcal{K}$, for all $A, B \in AR_K$,

$$(A,B) \in att(K) \text{ iff there is a defeasible } B' \in sub(B) \text{ such that } (A,B') \in att(K).$$

□

From left to right, (P5) reflects the accepted meaning of an attack for knowledge bases with defeasible and strict rules [27]. The right-to-left direction is commonly satisfied by attack relation assignments, with exceptions such as *lwl* [11].

PROPOSITION 5.5.
Subargument structure (P5) is satisfied by att_{swl}, att_{dwl} and att_{pdl}.

PROOF. **For** att_{swl}**.** ($\Rightarrow$) Let $(A,B) \in att_{swl}$(K). (Case: rebut.) Then, A contradicts some basic defeasible $B' \sqsubseteq B$ with $wl(A) \not< wl(B')$. Obviously such B' is a defeasible subargument of B. (Case: undercut.) In this case, there is $B' \sqsubseteq B$ such that $last(B') = d \in RD$ and $cnl(A) = ab_d$. It is again immediate that such B' is defeasible so that $(A,B') \in att_{swl}(K)$.

($\Leftarrow$) (Case: rebut.) If A contradicts a defeasible B' at a basic defeasible subargument 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 $(A,B) \in att_{swl}(K)$ is an attack at the same basic defeasible subargument $B'' \sqsubseteq B$. (Case: undercut.) If A undercuts B' at a basic defeasible subargument B'', then $cnl(A) = ab_d$ for $d = last(B'')$. Since $B'' \sqsubseteq B' \sqsubseteq B$, we conclude that A also undercuts B at a basic defeasible subargument B'' of B, and so $(A,B) \in att_{swl}(K)$.

For att_{dwl}**.** (Case: rebut.) The proof is analogous to that for att_{swl} except that we use $wl(A \setminus B') \not< wl(B' \setminus A)$ for ($\Rightarrow$), and $wl(B \setminus A) \leq wl(B' \setminus A) \leq wl(B'' \setminus A)$ for ($\Leftarrow$). (Case: undercut.) The previous proofs for both ($\Rightarrow$) and ($\Leftarrow$) also work for att_{dwl} since its definition of undercutting attack is the same as in att_{swl}.

For att_{pdl}**.** ($\Rightarrow$) This follows from Definition 16 and the fact that all strict arguments A are in all stable extensions of K. ($\Leftarrow$) Immediate from Definition 16. □

Attack closure (P6). This principle demands that all attacks are either undercuts[17] or contradicting arguments, and that all undercuts do count as attacks.

PRINCIPLE 6 (Attack closure). We say that att satisfies 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 5.6.
Attack closure (P6) is satisfied by att_{swl}, att_{dwl} and att_{pdl}.

PROOF. **For** att_{swl} **and** att_{dwl}**.** (1.) Attack closure can be seen to hold after a quick observation of Definitions 11 and 13–14, since rebuts are contradicting attacks.

(2.) Again, the definitions of att_{swl} and att_{dwl} (Defs. 13–14) show that all undercuts are always enforced within these attack relation assignments.

For att_{pdl}**.** Both claims (1)–(2) follow from Definition 16. (Recall also the observation after this definition). □

Effective rebut (P7). This principle 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 7 (Effective rebut). We say that att satisfies the effective rebut property 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 rebuts A_1, then

$$(A_0, A_1) \in att(K) \text{ iff } d_0 \nprec d_1.$$

□

PROPOSITION 5.7.
Effective rebut (P7) is satisfied by att_{swl} and att_{dwl}. It is not satisfied by att_{pdl}.

PROOF. **For** att_{swl}**.** Let $dr(A) = \{d_1\}$ and $dr(B) = \{d_2\}$ contain each one defeasible rule with A contradicting B at a basic defeasible $B' \sqsubseteq B$. Note that $wl(A) = rank(d_1)$ and $wl(B) = rank(d_2)$. ($\Rightarrow$.) Suppose that $(A, B) \in att_{swl}(K)$. Then, by definition, $wl(A) \nprec wl(B')$. Moreover, since B' is defeasible we must have $dr(B') = \{d_2\}$. Combining this with the inequality $wl(A) \nprec wl(B')$, we obtain $d_1 \nprec d_2$. ($\Leftarrow$.) Suppose now that $d_1 \nprec d_2$. So, $wl(A) \nprec wl(B)$. Since A contradicts B at a basic defeasible subargument $B' \sqsubseteq B$, we must have that $d_2 \in dr(B)$, and so $(A, B) \in att_{swl}(K)$.

For att_{dwl}**.** (Case $d_1 = d_2$.) Then clearly $d_1 \nprec d_2$ and also $(A, B) \in att_{dwl}(K)$. Hence, the equivalence follows. (Note moreover that in this case $cnl(A) = cnl(B)$ and each argument A, B attacks itself at a shared subargument C with $last(C) = d_1$).

(Case $d_1 \neq d_2$.) The proof is analogous to that for att_{swl}, once we take into account that $wl(dr(A) \setminus dr(B)) = wl(A)$ and $wl(dr(B) \setminus dr(A))) = wl(B)$.

For att_{pdl}**.** Recall the counterexample from the proof of (P1), with arguments:

$$A = [[\top] \overset{1}{\Rightarrow} a] \qquad B = [[\top] \overset{2}{\Rightarrow} \neg a] \qquad C = [[\top] \overset{3}{\Rightarrow} a].$$

[17] The notion of undercut from Principle 6 is the same as in Pollock [32] and ASPIC+ [28]: an argument A undercuts B at $B' \in sub(B)$ iff B's last rule is defeasible $d = last(B') \in RD$ and the attacking argument A states that this defeasible rule d is not applicable $cnl(A) = ab_d$.

The unique PDL extension is $E = \{a\}$, and so there is no PDL construction containing $\top \overset{2}{\Rightarrow} \neg a$. Thus, $(B, A) \notin att_{pdl}(K)$, despite the fact that $\top \overset{1}{\Rightarrow} a \prec \top \overset{2}{\Rightarrow} \neg a$. □

Link orientation (P8). The last of Dung's principles directs attacks against those links in an argument that are identified as responsible for this argument's weakness.

DEFINITION 19 (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 8 (Link orientation). Let $\mathcal{K}$ be a sensible class of knowledge bases and *att* be an attack relation assignment defined for $\mathcal{K}$. We say that *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

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

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

Dung motivates this principle as follows: *attacks should be directed against the culprit link within the attacked argument*. In our view, this phrasing conflates two notions of responsibility among the subarguments of an attacked argument: responsibility for (1) the conflict and (2) for the weakness of a link. While for *last link* (1) and (2) obviously always coincide, under *weakest link* (1) and (2) can apply to different subarguments. Take, for instance, C and resp. D in the counterexample found in Proposition 5.8 below. The weakness in (2) is clearly not preserved under the strengthening $C \longmapsto B$ in (P8), and so the property of (A, C) *being an attack* need not be preserved into the pair (A, B).

PROPOSITION 5.8.
Link orientation is not satisfied by any of the attacks att_{swl}, att_{dwl}. It is satisfied by att_{pdl}.

PROOF. A counterexample to (P8) for both att_{swl}, att_{dwl} can be found by expanding Example 1 with a new fact a. **For** att_{swl}**.** Let K consist of

$$RS = \emptyset \qquad RD = \{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\} \qquad BE = \{\top, a\}.$$

The set of arguments is $AR_K = \{[\top], [a], A, B, C, D\}$, where $A, \ldots, D$ are as follows:

$$\underbrace{[[\top] \Rightarrow \neg b]}_{A} \qquad \underbrace{[[a] \Rightarrow b]}_{B} \qquad \underbrace{[\overbrace{[\top \Rightarrow a]}^{D} \Rightarrow b]}_{C}$$

For $AS = \{D\}$, observe that argument C is a weakening of B, that is $C \in B \downarrow AS$. These arguments' strengths are $wl(A) = 2$, $wl(B) = 3$ and $wl(C) = wl(D) = 1$. Finally, observe that

$$(A, C) \in att_{swl}(K) \text{ and } (A, D) \notin att_{swl}(K) \quad \text{while} \quad (A, B) \notin att_{dwl}(K).$$

For att_{dwl}**.** The same example works, since for all the previous pairs $(X, Y) \in att_{swl}$, we have that $wl(X \setminus Y) = wl(X)$. Hence, $att_{dwl}(K) = att_{swl}(K)$ and the same violation exists: $(A, C) \in att_{dwl}(K)$ and $(A, D) \notin att_{dwl}(K)$ but $(A, B) \notin att_{swl}(K)$.

For att_{pdl}**.** Let $A, B, C \in AR_K$ with $C \in B \downarrow AS$ satisfy $(A, C) \in att_{pdl}(K)$ and $(A, X) \notin att_{pdl}(K)$ for all $X \in AS$. We prove that $(A, B) \in att_{pdl}(K)$. From $(A, C) \in att_{pdl}(K)$, there exists a PDL extension E of K, with construction R, such that (1) $dr(A) \subseteq R$, and A rebuts C at some C' satisfying $bd(last(C')) \subseteq E$. Since for all $X \in AS$, we have that $X \sqsubseteq C$ but $(A, X) \notin att_{pdl}(K)$, then A does not rebut any such X (at any subargument). Hence, (2) A rebuts B at some subargument B' with $C' \in B' \downarrow \{X\}$. Moreover, (3) $bd(last(B')) = bd(last(C')) \subseteq E$. Then, (1)–(3) jointly imply that $(A, B) \in att_{pdl}$. □

Weak context independence (P9). A principle that weakens (P2) was introduced in our previous work [11]. Below it is presented in a revised form. The new principle requires that the property (A, B) *is an attack* is invariant among all knowledge bases containing the same sub- and superarguments of A and B, and the same strengths in their defaults.[18] Let us define: $sup_K(A) = \{A^+ \in AR_K : A^+ \sqsupseteq A\}$.

PRINCIPLE 9 (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 $A, B \in AR_K \cap AR_{K'}$:

$$\text{if} \left\{ \begin{array}{c} sup_K(A) = sup_{K'}(A) \text{ and} \\ sup_K(B) = sup_{K'}(B) \text{ and} \\ \preceq, \preceq' \text{ agree on } dr(sup_K(A)) \cup dr(sup_K(B)) \end{array} \right\} \text{then} \quad \begin{array}{c} (A, B) \in att(K) \\ \text{if } \{f\} \\ (A, B) \in att(K'). \end{array}$$

□

PROPOSITION 5.9.
Weak context independence (P9) is satisfied by att_{swl}, att_{dwl}. It is not satisfied by att_{pdl}.

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 need to be tested for (P2): the former are all pairs validating Definition 9(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_{pdl}**.** The counterexample for (P2) works here as well, as no superargument plays a role in the facts that $(B, A) \in att_{pdl}(K')$ and $(B, A) \notin att_{pdl}(K)$. □

THEOREM 5.10.
The principles satisfied by each attack relation are listed in Table 1.

PROOF. These results for $att_{swl}, att_{dwl}, att_{pdl}$ follow from Propositions 5.1–5.9 above. For the case of att_{lwl}, we refer the reader to [11]. (Let us observe that the proofs in [11] are made under the assumption that $RS = \emptyset$. Those proofs can be easily expanded to the case with strict rules, as shown above for the attack relation assignments att_{swl}, att_{dwl}.) □

Discussion of the principle-based analysis. 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 based on some of the examples. On the other, the *prima facie* intuitive principles from Dung. Following Nelson Goodman

[18] The previous definition of (P9) did not require the strengths of defaults in superarguments to be the same. We are thankful to C. Strasser for pointing out a that att_{lwl} did not satisfy the old definition, and for suggesting the current definition.

[20]'s notion of *reflective equilibrium*, our analysis should prompt us to search for a balance between intuitions on principles and intuitions on cases. Let us take a look at Table 1.

(P1) Credulous cumulativity has also been challenged by Prakken and Vreeswijk [40, Section 4.4], and by Modgil and Prakken [29, Section 5.2]. Strengthening a defeasible conclusion may make it gain the ability to defeat more arguments, thereby causing a change in the stable extensions.

(P2) (Disjoint) weakest link verifies that attacks depend only on the defeasible rules of the two arguments in conflict, in contrast to PDL attacks.

(P3) Attack monotonicity is only violated by lookahead weakest link [11].

(P4)–(P6) Irrelevance of redundant defaults (P4) results in an intuitive property of ASPIC+, i.e. a semantic invariance under the weakening of facts into (irrelevant) defaults. Subargument structure (P5) also seems to capture a defining property of defeasible argumentation. Attack closure (P6) expresses our understanding of how attacks in ASPIC+ should be defined. These core principles (P4)–(P6) are agreed upon by most attacks considered in the literature.

(P7) Effective rebuts is satisfied by all attacks induced from common lifting functions: (disjoint) weakest link and last link. But argumentative approximations to PDL and its relational attacks are incompatible with it (Prop. 5.7).

(P8) Link orientation is also disputed: it clashes with all attack relation assignments inspired by weakest link, but not by PDL-based attacks.

(P9) Weak context independence was proposed to approximate PDL via attacks inspired by *lwl* in [11]. Surprisingly, it is not satisfied by PDL attacks, but it might still be worth considering for weakest link approximations to PDL.

In sum, while the principles proposed by Dung (P1)–(P8) offer an elegant presentation of attacks and their properties, these principles seem motivated by last link. Rather than foreclosing the existing debates on this question, we would like to open up the corresponding challenge for weakest link and its relatives, namely by searching for principles satisfied by the weakest link family, but not by last link. In this respect, the failure of (P3) and (P5) justifies our shift from lookahead weakest link att_{lwl} [11] to PDL attacks att_{pdl} (Definition 16). Still, att_{lwl} inspired a new principle (P9) to be explored within the weakest link family.

As a guide to what principles can be expected for an attack relation assignment *att* that captures exactly the PDL extensions, one might consider those principles satisfied by disjoint weakest link or pdl-attacks. These two attacks agree on (P3)–(P6) and disagree on (P2), (P7) and (P9).

It is conceivable though to validate all principles (P2)–(P9) under the following expansion of pdl-attacks (Definition 16).

DEFINITION 20 (PDL+ attack).
Given $K = (RS, RD, BE, rank)$, a knowledge subbase of K is any tuple $K' = (RS, RD', BE, \preceq')$ such that $RD' \subseteq RD$ and $\preceq'$ restricts $\preceq$ to RD'. Then, an argument A **PDL+ attacks** B, denoted $(A, B) \in att_{pdl+}(K)$, if there exists a knowledge subbase K' of K such that $(A, B) \in att_{pdl}(K')$.

This PDL+ attack relation assignment prevents the counterexample to the principles (P2), (P7), (P9) in Propositions 5.2, 5.7 and resp. 5.9.

EXAMPLE 10.
For $RD = \{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg a, \top \overset{3}{\Rightarrow} a\}$, let $RD' \subseteq RD$ consist of the first two defaults and define $K' = (RS, RD', rank, BE)$. Given the arguments $A = [[\top] \Rightarrow a]$ and $B = [[\top] \Rightarrow \neg a]$ in $AR_{K'}$, the

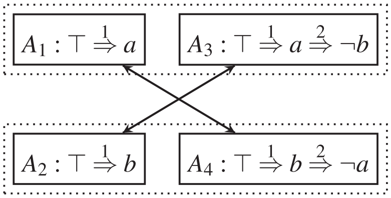

$A_1 : \top \overset{1}{\Rightarrow} a$
$A_3 : \top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$
$A_2 : \top \overset{1}{\Rightarrow} b$
$A_4 : \top \overset{1}{\Rightarrow} b \overset{2}{\Rightarrow} \neg a$

FIGURE 5. *AF* constructed from Example 11. Arrows describe all the possible individual attacks at subarguments. An attack relation $att(K)$ cannot contain both (A_1, A_4) and (A_2, A_3) if it is to capture exactly the two PDL extensions shown

PDL construcion $R = \{\top \Rightarrow \neg a\}$ gives the attack $(B, A) \in att_{pdl}(K')$, and then $(B, A) \in att_{pdl+}(K)$, as desired.

CONJECTURE 1.
The attack relation assignment att_{pdl+} satisfies (P2)–(P9) but its stable belief sets do not coincide with PDL extensions.

The impossibility theorem in the next section (Theorem 6.1) will provide evidence for this conjecture. This theorem implies that if att_{pdl+} satisfies (P2)–(P9), then some of its stable belief sets are not PDL extensions.

6 PDL and Dung's principles: an impossibility result

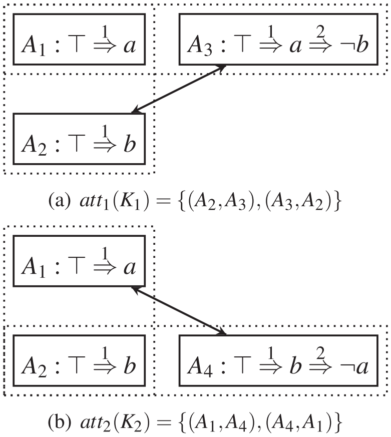
As previously discussed, most of the principles proposed by Dung [17] seem indisputable, yet some others hide a partisan view on what argumentation can or cannot be. Context independence, for example, could be used to rule out Brewka's PDL from argumentation altogether. For another example, credulous cumulativity is used by Dung [16, Example 7.1] directly against elitist orderings [27].

In this section, we offer more evidence against Context independence, in the form of an impossibility result (Theorem 6.1). Any attempt to realize PDL in ASPIC+ should preserve the definitional principle of Attack closure (P6). Theorem 6.1 explains how this is incompatible with Context independence (P2). Let us describe in three parts (Examples 11–13) the counterexample used in the proof.

EXAMPLE 11.
Define $RD = \{d_1 : \top \overset{1}{\Rightarrow} a, d_2 : \top \overset{1}{\Rightarrow} b, d_3 : a \overset{2}{\Rightarrow} \neg b, d_4 : b \overset{2}{\Rightarrow} \neg a\}$ as in the knowledge base $K = (RS, RD, \preceq, BE)$ from Example 3. Recall that PDL always selects a strongest applicable and consistent default. This gives the outputs:

$$\begin{aligned} R_1 = \{d_1, d_3\} &\quad\longmapsto\quad S_1 = \{a, \neg b\} \\ R_2 = \{d_2, d_4\} &\quad\longmapsto\quad S_2 = \{b, \neg a\}. \end{aligned}$$

Figure 5 shows an AF for K with all rebuts displayed. (We omit $\top$ from the PDL extensions and the argument $A_0 = [\top]$ from AR_K).


$A_1 : \top \overset{1}{\Rightarrow} a$
$A_3 : \top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$
$A_2 : \top \overset{1}{\Rightarrow} b$
$A_1 : \top \overset{1}{\Rightarrow} a$
$A_2 : \top \overset{1}{\Rightarrow} b$
$A_4 : \top \overset{1}{\Rightarrow} b \overset{2}{\Rightarrow} \neg a$

(a) $att_1(K_1) = \{(A_2,A_3),(A_3,A_2)\}$

(b) $att_2(K_2) = \{(A_1,A_4),(A_4,A_1)\}$

FIGURE 6. (a) Under the attack relation att_1, the two stable extensions of $AF_1 = (AR_{K_1}, att_1(K_1))$ match the two PDL extensions of K_1 in Example 12. (b) The stable extensions under att_2 similarly match the PDL extensions of K_2 in Example 13

EXAMPLE 12.
Let $K_1 = (RS, RD_1, \preceq_1, BE)$ be the fragment of K consisting of $RD_1 = RD \setminus \{d_4\}$ and $\preceq_1$ being the restriction of $\preceq$ to the set RD_1. PDL then outputs:

$$\begin{array}{lcl} R_1 = \{d_1, d_3\} & \longmapsto & S_1 = \{a, \neg b\} \\ R_3 = \{d_2, d_1\} & \longmapsto & S_3 = \{a, b\}. \end{array}$$

The PDL extensions S_1, S_3 also obtain as the stable belief sets under the attack relation $att_1 = \{(A_2,A_3),(A_3,A_2)\}$. See Figure 6(a) for an illustration.

EXAMPLE 13.
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

$$\begin{array}{lcl} R_2 = \{d_2, d_4\} & \longmapsto & S_2 = \{b, \neg a\} \\ R_3 = \{d_1, d_2\} & \longmapsto & S_3 = \{a, b\}. \end{array}$$

The PDL extensions S_2, S_3 again obtain from stable extensions, now under the attack relation $att_2 = \{(A_1,A_4),(A_4,A_1)\}$. See Figure 6(b).

Now we are in a position to prove the 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 [2].

THEOREM 6.1.
Let *att* be an attack relation assignment capturing the PDL extensions (say, under stable semantics) and satisfying attack closure (P6). 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 11–13. Let also *att* be an attack relation assignment capturing the PDL extensions under stable semantics.

Given this attack relation assignment *att*, the stable extensions must be the following:

$AF_0 = (AR_K, att(K))$	$AF_1 = (AR_{K_1}, att(K_1))$	$AF_2 = (AR_{K_2}, att(K_2))$
$\mathcal{E}_1 = \{A_1, A_3\}$	$\mathcal{E}_1 = \{A_1, A_3\}$	
$\mathcal{E}_2 = \{A_2, A_4\}$		$\mathcal{E}_2 = \{A_2, A_4\}$
	$\mathcal{E}_3 = \{A_1, A_2\}$	$\mathcal{E}_3 = \{A_1, A_2\}$

The proof is by contradiction. Assume context independence (P2). Using attack closure (P6), 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 (P2), 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 become a stable extension of $AF_0 = (AR_K, att(K))$ without being a PDL extension of K. Hence, context independence is not satisfied. □

Some readers might wonder if Theorem 6.1 can be read as providing evidence against PDL rather than Context independence (P2). To this, one might reply by acknowledging that this axiom conveniently simplifies (the study of) attack relations, but that our intuitions on these relations are not yet strong enough to blindly embrace all the consequences of (P2). For now, we leave open this dilemma between PDL and (P2).

7 Approximating DWL from a variant of PDL.

As seen in Examples 7–8, disjoint weakest link and PDL are incomparable under total preorders. As a first step towards their convergence, one can slightly modify PDL to make it closer to disjoint weakest link. To this end, we propose *parallel* PDL (pPDL), a concurrent variant of PDL. The main novelty of pPDL is that each inductive step can concurrently select a set of defaults, rather than just one.

DEFINITION 21 (pPDL).
Let $K = (RS, RD, \preceq, BE)$ be a knowledge base. A *parallel PDL extension* of K, also called **pPDL extension**, is any set $E = \bigcup_n E_n$ inductively defined by

$$E_0 = BE$$

$$E_{n+1} = \begin{cases} E_n \cup hd[D_{n+1}] & \text{if a set } D_{n+1} \text{exists satisfying (1)–(2) below} \\ E_n & \text{otherwise} \end{cases}$$

where $D_{n+1} = \{r_1, \ldots, r_k\}$ is a non-empty set of defaults that are (1) $\preceq$-maximal among the active defaults in E_n, and such that (2) $Cl_{RS}(E_n \cup hd[D_{n+1}])$ is non-contradictory. The sequence $R = (D_n)_{n<\omega}$ is called a *pPDL construction* for K.

FACT 7.1. Let K be a prioritized default theory. Each PDL extension of K is also a pPDL extension of K. □

PROOF. A pPDL selection $R = (D_n)_{n<\omega}$ for K can, in particular, simply choose a singleton set $D_{n+1} = \{r_{n+1}\}$ at each inductive step $E_n \mapsto E_{n+1}$. Observe that in this case conditions (1)–(2) from Definition 21 collapse into condition ($\star$) for the inductive step of a PDL extension (Definition 15), their only difference being in the form of the corresponding constructions: $(\{r_1\}, \{r_2\}, \ldots)$ and resp. $(r_1, r_2, \ldots)$. □

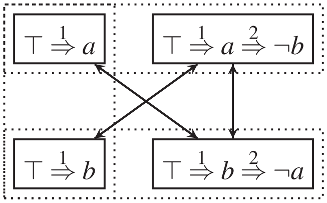

$\top \overset{1}{\Rightarrow} a$
$\top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$
$\top \overset{1}{\Rightarrow} b$
$\top \overset{1}{\Rightarrow} b \overset{2}{\Rightarrow} \neg a$

FIGURE 7. pPDL differs from PDL. PDL has two extensions $\{a, \neg b\}$ and $\{b, \neg a\}$. pPDL has an additional extension $\{a, b\}$. Arrows denote logical conflicts

EXAMPLE 14 (pPDL, DWL *vs* PDL).
Let us use Example 3 to show that the default logic PDL differs from pPDL. Figure 7 illustrates the three pPDL extensions $\{a, \neg b\}$, $\{b, \neg a\}$, $\{a, b\}$, of which $\{a, b\}$ is not a PDL extension.

In the impossibility theorem (Theorem 6.1), Context independence (P2) requires that the extensions of K (Figure 7) are the sum of the extensions of K_1 and K_2 (Figures 6(a)–6(b)). Indeed for pPDL, att_{swl} and att_{dwl}, these are of the form:

$$\big\{\{a, \neg b\}, \{b, \neg a\}, \{a, b\}\big\} = \big\{\{a, \neg b\}, \{a, b\}\big\} \cup \big\{\{b, \neg a\}, \{a, b\}\big\}.$$

Although pPDL and att_{dwl} agree on this and other examples, pPDL does not always match disjoint weakest link.

EXAMPLE 15 (pPDL *vs* DWL).
Example 8 showed a unique stable belief set $\{a, b, c\}$ under att_{dwl}. There are though two pPDL extensions: $\{a, b, c\}$ and $\{a, \neg b, c\}$.

A summary of the results in Sections 3 and 7 is shown in Table 2. The remaining potential inclusions (labelled '?') are unknown, but conjectured to be true.

8 Related work

There is a lot of work in the nonmonotonic logic and logic programming literature on prioritized rules; see e.g. Delgrande et al. [12] for an overview. A long-standing problem in these areas, to reduce the large number of extensions of a knowledge base and favour those with intuitive outcomes, has been addressed with justifications, undercutters or priorities—raising in turn further questions on the semantics of argumentation and prioritized logics. Prakken and Horty [38] explain that Pollock [34, 35] used weakest link to define argument strength—inspired, in fact, by probabilistic considerations. Beirlaen et al. [3] also point out that weakest link is defined purely in terms of the strength of the defeasible rules used in argument construction. Pardo and Straßer [30] give an overview of argumentative representations of prioritized default logics, mainly using disjoint weakest link *dwl*. Our goal is also to explain prioritized rule-based systems with argumentation semantics.

Various authors have discussed the dilemma between weakest link and last link [9, 24, 27, 28]. The analysis of simple weakest link *swl* indicates that the general idea of weakest link is more complicated and ambiguous than it seems at first sight. With partial orders, ASPIC+ combines weakest link and last link either with an elitist ordering (as in the present work) or with a democrating ordering. Discussions of the dilemma, though, center around the elitist ordering, with inconclusive results. For **descriptive** applications, as in Example 1 (the fitness-loving Scot), Modgil and Prakken [28, 29] opt for weakest link as it correctly propagates the uncertainty (of a weakest link, in particular) into superarguments. Dung appeals to defence of last link is based on credulous cumulativity: in the dean sceanrio (Example 5), upgrading a previous inference (*professor*) into a fact should preserve other conclusions as well. This is the case for last link (w.r.t. *teach*) but not for weakest link (¬*teach* turns into *teach* after the upgrade).

Young et al. [42, 43] show that even for total and modular orders, *swl* cannot always give intuitive conclusions. They also show the correspondence between the inferences made in PDL and *dwl* with strict total orders. Then they raise the question of the similarity between weakest link and PDL for modular orders (that we address) and for partial orders. Moreover, Liao et al. [24] give similar results but use other examples to demonstrate that the approach of Young et al. [42, 43] cannot be extended to preorders [24]. Liao et al. [24] use an order puzzle in the form of Example 3 to show that even with modular orders, selecting the correct reasoning procedure is challenging. Finally, Lehtonen et al. present novel complexity results for ASPIC+ with preferences that are based on weakest link (*swl* in this paper) [23], by rephrasing the stable semantics in terms of subsets of defeasible elements. While these articles provide technical insights, they are not decisive for either solution of the dilemma—thus prompting the axiomatic study of Dung [16].

Priorities, often read in terms of normalcy or typicality in epistemic scenarios, have become also popular in normative systems. Constitutive norms, in combination with brute facts, use priorities for deciding about legal or institutional facts (e.g. that the snoring professor *misbehaves* in Example 1). Deontic norms, which are often conditional on institutional facts, use priorities to capture natural hierarchies of norms, based on authority, recency or specificity. For **normative** applications, the dilemma between weakest and last link is also discussed. Modgil and Prakken [28, 29] lean into last link—in sharp contrast to the descriptive case. Their reasoning is that while obligations build on facts or claims, a normative conflict between two obligations should not be decided at all on the basis of the strengths of these facts or claims—whatever these strengths are, they suffice anyway for these claims to be accepted.

While compelling, this argument does not apply to all deontic logic or argumentation systems. First, for systems without descriptive defaults (including constitutive rules), all discussions are normative and so one may freely apply weakest link, if desired. Hence, last link is not the only option for deontic applications. The inference rules of a deontic system, in other cases, may collapse the two choices in this dilemma. This is the case for systems that do not allow deontic detachment, i.e. the chaining of norms, so that only facts (not obligations) can trigger an obligation. Normative claims are then of depth one and the distinction between weakest and last link collapses. Against the above argument for last link are also systems that establish a preference for factual arguments over obligation arguments. Such a preference, used to prevent wishful thinking, can be found in the work of Pigozzi and van der Torre [31] and in its application to machine ethics by Liao et al. [25] for the regulation of an agent by multiple stakeholders.

The argument for last link in deontic applications finally depends on how we interpret the strengths of descriptive defaults. While the natural reading for norms is that of authority or importance, for ordinary defaults strength is ultimately read in terms of uncertainty. In this case, factual claims (particularly, those with weak evidential support) are very relevant to the discussion. Suppose that Rob was slapped by a zombie, and this fact gives weak support for his upcoming infection:

$$RD = \left\{ \begin{array}{c} \mathit{ZSlaps}(\mathrm{Rob}) \overset{1}{\Rightarrow} \mathit{Infected}(\mathrm{Rob}),\ \mathit{Infected}(\mathrm{Rob}) \overset{1}{\Rightarrow\mkern-14mu\blacktriangleright} \mathit{LockUp}(\mathrm{Rob}), \\ \top \overset{2}{\Rightarrow\mkern-14mu\blacktriangleright} \neg\mathit{Kill}(\mathrm{Rob}),\ \mathit{Infected}(\mathrm{Rob}) \overset{3}{\Rightarrow\mkern-14mu\blacktriangleright} \mathit{Kill}(\mathrm{Rob}) \end{array} \right\}$$

Say a default $p \overset{1}{\Rightarrow} q$ means a conditional probability $Pr(q|p) \geq .1$, while nothing else is known about the conditional for non-infection, i.e. $Pr(\neg q|p) \leq .9$. The rest of *RD* describes, using $\Rightarrow\!\!\!\blacktriangleright$, an unconditional norm and a protocol for zombie infections. Let also $BE = \{ZSlaps(\text{Rob})\}$. Then, the weakest link output *to lock up Rob and not kill him* seems more intuitive than *killing Rob*, based on last link.

In sum, even if one settles the dilemma for the descriptive case with axioms or principles, further considerations might still be needed in deontic applications in order to decide between weakest and last link.

9 Summary and future work

In this article, we advanced on the study of weakest link by comparing logic and argumentation-based realizations of this idea, mainly in the form of Brewka's PDL and (disjoint) weakest link. Our ultimate goal is two-fold: to bridge the gap between these two areas; and to contribute to the debate about weakest and last link, a central dilemma in defeasible reasoning and argumentation. To this end, we adopted the formal framework of attack relation assignments proposed by Dung [16].

The first goal led us to identify an attack relation that captures PDL extensions, and to compare it with attacks based on simple and disjoint weakest link using the eight principles (P1)–(P8) advanced by Dung. We proved which principles for attack relations are satisfied by weakest link, disjoint weakest link and PDL-based attacks. In this respect, we extended earlier results [11] to strict rules and confirmed our conjecture on this question. Our principle-based analysis (Table 1) has several original insights, presents the difference between several kinds of attack relation assignment, identifies and explains the nature of the weakest link principle and reveals there is still some potential for weakest link attack to improve. On this last question, we proposed pPDL (parallel PDL), a concurrent variant of PDL and showed by way of examples that it falls closer to disjoint weakest link than PDL does. While this pPDL variant still does not match disjoint weakest link, one might conjecture that some further refinement might do.

For the second goal, the principle-based analyses mentioned above 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. In particular, these analyses can be used to tell apart those principles that express common aspects of argumentation, such as (P3)–(P6), from those that are uncertain (P7)–(P9) and finally those that have been challenged (P1)–(P2) by different authors. An impossibility theorem between two of these principles (P2) and (P6) in the context of PDL reinforced this view on our evaluation of all these principles.

The core idea behind weakest link is, in our opinion, at least as important as last link for general applications in AI. Our results thus motivate a search for other principles for weakest link that are not satisfied by last link. A fine-grained characterization of a class of weakest link attack relation assignments, in the style of the characterizations proposed by Dung for last link [16, 18], would also help us deepen our understanding of weakest link and vindicate its use in argumentation and non-monotonic reasoning. As a step in this direction, we also studied a new principle *weak context independence* (P9). This principle is a weakening of context independence (P2) that can be satisfied by variants of weakest link that violate (P2), as shown for *lwl* (lookahead weakest link) in previous work [11].

For future work, different questions remain open. First, a major challenge is how to generalize recent insights from this article and the related literature to partial orders as studied in ASPIC+. From a representation point of view, total orders give only one extension, while partial or modular orders may produce multiple extensions. Another open question for the future of ASPIC+-style structured argumentation is which way to go: introduce auxiliary arguments like Liao et al. [24], or weaken context independence. 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 case. We thus leave for future work deciding whether this is the case for the attack relation assignments under the present study, and for other alternatives also inspired by weakest link yet to be considered. Related to this question is that of identifying a variant of PDL that coincides with disjoint weakest link. 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 would like to thank two anonymous reviewers and Emil Weydert for their helpful suggestions and remarks.

References

[1] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, **29**, 125–169, 2002.

[2] K. J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, **58**, 328–346, 1950.

[3] M. Beirlaen, J. Heyninck, P. Pardo and C. Straßer. Argument strength in formal argumentation. *Journal of Logics and their Applications - IfCoLog*, **5**, 629–676, 2018.

[4] P. Besnard and A. Hunter. *Elements of Argumentation*. The MIT Press, Cambridge, MA, 2008.

[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. Reasoning about priorities in default logic. In *Proc. of the 12th National Conference on AI*, B. Hayes-Roth and R. E. Korf, eds, vol. **2**, pp. 940–945. AAAI Press/The MIT Press, Cambridge, MA, 1994.

[7] G. Brewka and T. Eiter. Preferred answer sets for extented logic programs. *Artificial Intelligence*, **109**, 297–356, 1999.

[8] G. Brewka and T. Eiter. Prioritizing default logic. In *Intellectics and Computational Logic*, S. Hölldobler, ed., vol. **19** of *Applied Logic Series*, pp. 27–45. Springer Netherlands, Dordrecht, 2000.

[9] M. Caminada. Rationality postulates: Applying argumentation theory for non-monotonic reasoning. *Journal of Applied Logics - IfCoLog*, **4**, 2457–2492, 2017.

[10] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, **171**, 286–310, 2007.

[11] C. Chen, P. Pardo, L. van der Torre and L. Yu. Weakest link in formal argumentation: Lookahead and principle-based analysis. In *Proc. of Logic and Argumentation - 5th International Conference, CLAR 2023, volume 14156 of Lecture Notes in Computer Science*, A. Herzig, J. Luo and P. Pardo, eds, pp. 61–83. Springer, Berlin Heidelberg, 2023.

[12] J. P. Delgrande and T. Schaub. Expressing preferences in default logic. *Artificial Intelligence*, **123**, 41–87, 2000.

[13] J. P. Delgrande, T. Schaub, H. Tompits and K. Wang. A classification and survey of preference handling approaches in nonmonotonic reasoning. *Computational Intelligence*, **20**, 308–334, 2004.

[14] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, **77**, 321–357, 1995.

[15] P. M. Dung. An axiomatic analysis of structured argumentation for prioritized default reasoning. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, volume 263 of Frontiers in Artificial Intelligence and Applications*, T. Schaub, G. Friedrich and B. O'Sullivan, eds, pp. 267–272. IOS Press, Amsterdam, 2014.

[16] P. M. Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, **231**, 107–150, 2016.

[17] P. M. Dung. A canonical semantics for structured argumentation with priorities. In *Computational Models of Argument - Proceedings of COMMA, volume 287 of Frontiers in Artificial Intelligence and Applications*, P. Baroni, T. F. Gordon, T. Scheffler and M. Stede, eds, pp. 263–274. IOS Press, Amsterdam, 2016.

[18] P. M. Dung and P. M. Thang. Fundamental properties of attack relations in structured argumentation with priorities. *Artificial Intelligence*, **255**, 1–42, 2018.

[19] P. M. Dung, P. M. Thang and T. C. Son. On structured argumentation with conditional preferences. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI*, pp. 2792–2800. AAAI Press, Cambridge, MA, 2019.

[20] N. Goodman. *Fact, Fiction, and Forecast*. Harvard University Press, Cambridge, MA, 1955.

[21] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical logic arguments: postulates and properties. *Artificial Intelligence*, **175**, 1479–1497, 2011.

[22] G. Governatori, M. J. Maher, G. Antoniou and D. Billington. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, **14**, 675–702, 2004.

[23] Tuomo Lehtonen, Johannes P Wallner and Matti Järvisalo. 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*, Luc de Raedt, ed, vol **19**, pp. 215–225. IJCAI Organization, Menlo Park, CA, 2022.

[24] B. Liao, N. Oren, L. van der Torre and S. Villata. Prioritized norms in formal argumentation. *Journal of Logic and Computation*, **29**, 215–240, 2019.

[25] B. Liao, P. Pardo, M. Slavkovik and L. van der Torre. The Jiminy Advisor: moral agreements among stakeholders based on norms and argumentation. *Journal of Artificial Intelligence Research*, **77**, 737–792, 2023.

[26] D. Makinson. General patterns in nonmonotonic reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, H. Gabbay and P. Robinson, eds, vol. **3**, pp. 35–110. Oxford University Press, Oxford, 1994.

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

[28] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, **5**, 31–62, 2014.

[29] S. Modgil and H. Prakken. Abstract rule-based argumentation. In *Handbook of Formal Argumentation*, , et al. S. Baroni, eds, vol. **1**, pp. 287–364. College Publications, 2018.

[30] P. Pardo and C. Straßer. Modular orders on defaults in formal argumentation. *Journal of Logic and Computation* (online: exac084), Rickmansworth, 2022.

[31] G. Pigozzi and L. van der Torre. Arguing about constitutive and regulative norms. *Journal of Applied Non-Classical Logics*, **28**, 189–217. Oxford, 2018.

[32] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, **11**, 481–518, 1987.

[33] J. L. Pollock. How to reason defeasibly. *Artificial Intelligence*, **57**, 1–42, 1992.

[34] J. L. Pollock. Justification and defeat. *Artificial Intelligence*, **67**, 377–407, 1994.

[35] J. L. Pollock. *Cognitive Carpentry: A Blueprint for how to Build a Person*. The MIT Press, Cambridge, MA, 1995.

[36] J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, **133**, 233–282, 2001.

[37] J. L. Pollock. Defeasible reasoning and degrees of justification. *Argument & Computation*, **1**, 7–22, 2010.

[38] H. Prakken and J. F. Horty. An appreciation of John Pollock's work on the computational study of argument. *Argument & Computation*, **3**, 1–19, 2012.

[39] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, **7**, 25–75, 1997.

[40] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In *Handbook of Philosophical Logic*, pp. 219–318. Springer Netherlands, Dordrecht, 2002.

[41] T. Reid. *Essays on the Intellectual Powers of Man*. Cambridge University Press, Cambridge, 2011.

[42] A. P. Young, S. Modgil and O. Rodrigues. Prioritised default logic as rational argumentation. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, C. M. Jonker, S. Marsella, J. Thangarajah and K. Tuyls, eds, pp. 626–634. ACM, New York, 2016.

[43] A. P. Young, S. Modgil and O. Rodrigues. On the interaction between logic and preference in structured argumentation. In *Theory and Applications of Formal Argumentation - 4th International Workshop*, E. Black, S. Modgil and N. Oren, eds, vol. **10757**, pp. 35–50. Springer, Cham, Berlin Heidelberg, 2017.

Received 30 April 2024