2 Basic Notions and Notation

2.1 Specifying Rules and their Theories

For the remainder of this paper, let us narrow the general logical setting of specificity down to the concrete framework of *defeasible logic with the restrictions of positive-conditional specification with an inactive negation symbol*, as found e.g. in [STOLZENBURG &AL., 2003] and [CHESÑEVAR &AL., 2003].

In effect, these restrictions give us the standard "definite rules" of positive-conditional specification (or HORN-clause logic). Positive-conditional specification differs from logic programming in PROLOG (cf. e.g. [KOWALSKI, 1974], [CLOCKSIN & MELLISH, 2003]) insofar as termination issues and the order of the definite clauses are irrelevant for the semantics, and insofar as there is no cut predicate(s) and no negation as failure.

Such *definite rules* are implications of the following form: The conclusion is an atom; the condition is a (possibly empty) conjunction of (positive) atoms which may contain extra variables (i.e. free variables not occurring in the conclusion). This is can be seen as quantifier-free first-order logic with specifications restricted to implications of the mentioned form.

We ask the reader not to get confused on the mentioned effective form of our rules by the fact that — in place of the atoms — literals resulting from an inactive negation symbol are actually admitted in the rules of Definition 2.1. This special form of negation is standard in defeasible logic for convenience in the application context (such as an argumentation framework). In this paper, however, we can consider this negation just as a form of syntactical sugar (cf. Definition 2.2, Remark 2.3).

Definition 2.1 (Literal, Rule)
A *term* is inductively defined to be either a symbol for a free variable or a function symbol applied to a (possibly empty) list of terms. In the following, we will mainly use constants, e.g. tweety, which are function symbols without arguments.
An *atom* consists of a predicate symbol applied to a (possibly empty) list of terms, e.g. bird(tweety), meaning that tweety is a bird.
A *literal* is an atom, possibly prefixed with the symbol "¬" for negation.
A *rule* is a literal, but possibly suffixed with a reverse implication symbol "⇐" that is followed by a conjunction of literals, consisting of one literal at least.

Definition 2.2 (Theory, Derivation)
Let Π be a set of rules. The *theory of* Π is the set $\mathfrak{T}_\Pi$ inductively defined to contain

- all instances of literals from Π and
- all literals L for which there is a conjunction C of literals from $\mathfrak{T}_\Pi$ such that $L \Leftarrow C$ is an instance of a rule in Π.

For $\mathfrak{L} \subseteq \mathfrak{T}_\Pi$, we also say that Π *derives* $\mathfrak{L}$, and write $\Pi \vdash \mathfrak{L}$.

2.3 Global Parameters for the Given Specification

Throughout this paper, we will assume a set of literals Π^{F} and two sets of rules Π^{G}, Δ to be given:

- A set Π^{F} of literals meant to describe the *facts* of the concrete situation under consideration,
- a set Π^{G} of *general rules* $L \in C$ meant to hold in all possible worlds,[3] and
- a set Δ of *defeasible* (or default) rules meant to hold in most situations.

The set $\Pi := \Pi^{\mathrm{F}} \cup \Pi^{\mathrm{G}}$ is the set of *strict* rules that — contrary to the defeasible rules — are considered to be safe and are not doubted in the concrete situation.

2.4 Formalization of Arguments

? There is no difference in derivation between the strict rules from Π and the defeasible rules from Δ. If a contradiction occurs, however, we will narrow the defeasible rules from Δ down to a subset $\mathcal{A}$ of its *ground* instances (i.e. instances without free variables) — such that no further instantiation can occur. Such a subset, together with the literal whose derivation is in focus, is called an *argument*. With implicit reference to the given sets of rules Π and Δ, the formal definition is as simple as follows.

Definition 2.6 ([Contradictory] [Minimal] Argument)
? $(\mathcal{A}, L)$ is an [*minimal*] *argument* if $\mathcal{A}$ is a set of ground instances of rules from Δ and $\mathcal{A} \cup \Pi \vdash \{L\}$ [and $\mathcal{A}' \cup \Pi \nvdash \{L\}$ for any proper subset $\mathcal{A}' \subsetneq \mathcal{A}$].
An argument $(\mathcal{A}, L)$ is *contradictory* if $\mathcal{A} \cup \Pi$ is a contradictory set of rules.

[The reader may not understand the meaning of the brackets here.

Remark 2.7 (Non-Ground Arguments)
? What we ~~would~~ actually need is not exactly a set $\mathcal{A}$ of *ground* instances, but just of the instances appearing in the derivation. Then, however, we have to freeze the variables in $\mathcal{A}$ because they must not be instantiated in the derivation $\mathcal{A} \cup \Pi \vdash \{L\}$. We avoid this
? novel complication here ~~at first~~ because it does not play an essential rôle before § 8.3.

[3] In the approach of [Stolzenburg &al., 2003], the set Π^{G} must not contain mere literals (without suffixed condition). To obtain a more general setting, we omit this additional restriction in the context of this paper, simply because it is neither intuitive nor required for our framework here. For the actual occurrence of a literal in Π^{G}, see the discussion of Example 7.7 in § 7.4.

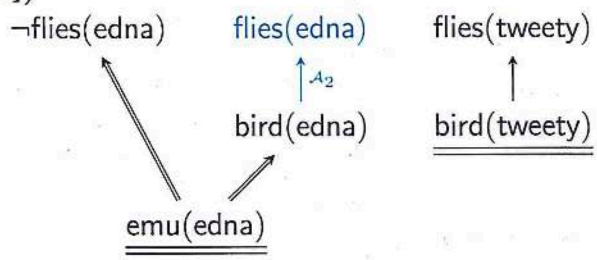
3 Motivating ~~Toy~~ Examples [superfluous to mention]

For ease of distinction, we will use the special symbol "$\leftarrow$" as a syntactical sugar in concrete examples of defeasible rules from Δ, instead of the symbol "$\Leftarrow$", which — in our concrete examples — will be used only in strict rules.

Moreover, in our graphical illustrations we will indicate membership in Π^{F} by *double underlining*. [paragraphs should not consist of only one sentence]

Example 3.1 (Example 1 of [Poole, 1985])

$$
\begin{array}{lcl}
\Pi^{\mathrm{F}}_{3.1} & := & \left\{ \begin{array}{l} \mathsf{bird}(\mathsf{tweety}), \\ \mathsf{emu}(\mathsf{edna}) \end{array} \right\}, \\
\Pi^{\mathrm{G}}_{3.1} & := & \left\{ \begin{array}{l} \mathsf{bird}(x) \Leftarrow \mathsf{emu}(x), \\ \neg\mathsf{flies}(x) \Leftarrow \mathsf{emu}(x) \end{array} \right\}, \\
\Delta_{3.1} & := & \left\{ \ \mathsf{flies}(x) \leftarrow \mathsf{bird}(x) \ \right\}, \\
\mathcal{A}_2 & := & \left\{ \ \mathsf{flies}(\mathsf{edna}) \leftarrow \mathsf{bird}(\mathsf{edna}) \ \right\}.
\end{array}
$$



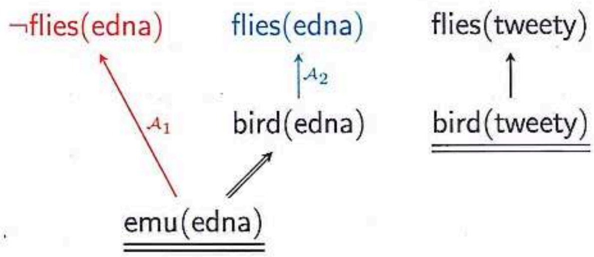
We have
$$
\begin{array}{lcl}
\mathfrak{T}_{\Pi_{3.1}} & = & \{\mathsf{bird}(\mathsf{tweety}), \mathsf{emu}(\mathsf{edna}), \mathsf{bird}(\mathsf{edna}), \neg\mathsf{flies}(\mathsf{edna})\}, \\
\mathfrak{T}_{\Pi_{3.1}\cup\Delta_{3.1}} & = & \{\mathsf{flies}(\mathsf{edna}), \mathsf{flies}(\mathsf{tweety})\} \cup \mathfrak{T}_{\Pi_{3.1}}.
\end{array}
$$

It is intuitively clear that we prefer the argument $(\emptyset, \neg\mathsf{flies}(\mathsf{edna}))$ to the argument $(\mathcal{A}_2, \mathsf{flies}(\mathsf{edna}))$, simply because the former does not use any defeasible rules. We will further discuss this in Example 6.17.

Let us see what happens to Example 3.1 if, ~~returning home from a tent show with pigs riding bicycles,~~ we are not so certain anymore that no emu can fly and turn the general rule $(\neg\mathsf{flies}(x) \Leftarrow \mathsf{emu}(x)) \in \Pi^{\mathrm{G}}_{3.1}$ into a defeasible one in the following example.

Example 3.2 (Example 2 of [Poole, 1985])

$$
\begin{array}{lcl}
\Pi^{\mathrm{F}}_{3.2} & := & \left\{ \begin{array}{l} \mathsf{bird}(\mathsf{tweety}), \\ \mathsf{emu}(\mathsf{edna}) \end{array} \right\}, \\
\Pi^{\mathrm{G}}_{3.2} & := & \left\{ \ \mathsf{bird}(x) \Leftarrow \mathsf{emu}(x) \ \right\}, \\
\Delta_{3.2} & := & \left\{ \begin{array}{l} \neg\mathsf{flies}(x) \leftarrow \mathsf{emu}(x), \\ \mathsf{flies}(x) \leftarrow \mathsf{bird}(x) \end{array} \right\}. \\
\mathcal{A}_1 & := & \left\{ \ \neg\mathsf{flies}(\mathsf{edna}) \leftarrow \mathsf{emu}(\mathsf{edna}) \ \right\}. \\
\mathcal{A}_2 & := & \left\{ \ \mathsf{flies}(\mathsf{edna}) \leftarrow \mathsf{bird}(\mathsf{edna}) \ \right\}.
\end{array}
$$



We have
$$
\begin{array}{lcl}
\mathfrak{T}_{\Pi_{3.2}} & = & \{\mathsf{bird}(\mathsf{tweety}), \mathsf{emu}(\mathsf{edna}), \mathsf{bird}(\mathsf{edna})\}, \\
\mathfrak{T}_{\Pi_{3.2}\cup\Delta_{3.2}} & = & \{\neg\mathsf{flies}(\mathsf{edna}), \mathsf{flies}(\mathsf{edna}), \mathsf{flies}(\mathsf{tweety})\} \cup \mathfrak{T}_{\Pi_{3.2}}.
\end{array}
$$

It is intuitively clear that we prefer the argument $(\mathcal{A}_1, \neg\mathsf{flies}(\mathsf{edna}))$ to the argument $(\mathcal{A}_2, \mathsf{flies}(\mathsf{edna}))$, simply because the defeasible derivation of the former is based on $\mathsf{emu}(\mathsf{edna})$, and because this is more specific than $\mathsf{bird}(\mathsf{edna})$, on which the derivation of the latter argument is based. We will further discuss this in Example 6.19.

4 Toward an Intuitive Notion of Specificity

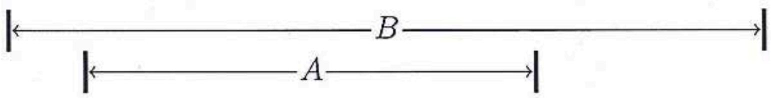
4.1 The Common-Sense Concept of Specificity

[always with or without hyphen, I prefer the latter]

It is part of general knowledge that a criterion is [properly] more specific than another one if the "class of candidates that satisfy it" is a [proper] subclass of that of the other one.

Analogously — taking logical formulas as the criteria — a formula A is [properly] more specific than a formula B, if the model class of A is a [proper] subclass of the model class of B, i.e. if $A \models B$ [and $B \not\models A$].

If we consider a formula as a predicate on model-theoretic structures, its model class becomes the extension of this predicate. From this viewpoint, we can state $A \models B$ also as the syllogism "every A is B", and also as the LAMBERT diagram[4] [no footnote, but inline is better]



4.2 Arguments as an Intuitive Abstraction

To enable a closer investigation of the critical parts of a defeasible derivation, we have to isolate the defeasible parts in the derivation. From a concrete derivation of a literal L, let us abstract the set $\mathcal{A}$ of the ground instances of the defeasible rules that are actually applied in the derivation, and form the pair $(\mathcal{A}, L)$, which we already called an *argument* in Definition 2.6 of § 2.4.

[4] Cf. [LAMBERT, 1764, Dianoiologie, §§ 173–194]

We have to exclude Π^F from this comparison, however. This exclusion makes sense because the defeasible rules are typically default rules not written in particular for the given concrete situation that is formalized by Π^F. Moreover, as indicated before, the inclusion of Π^F would typically eliminate all differences between activation sets, such as it is the case in all examples of § 3.

Finally, as we want to compare the defeasible parts of derivations, we should exclude the set Δ of the defeasible rules when we compare activation sets.

Thus, on the one hand, all we can take into account from our specification is a subset of the general rules Π^G, and, on the other hand, we do not want to exclude any of these general rules.

All in all, we conclude that Π^G is that part of our specification modulo which activation sets are to be compared.

4.3.2 A first Sketch of a Notion of Specificity

Very roughly speaking, if we have fewer activation sets for the defeasible part of a derivation, then these activation sets describe fewer models (i.e. their disjunction has fewer models), which again means that the defeasible part of the derivation is more specific. Accordingly, a first sketch of a notion of specificity can now be given as follows:

> An argument $(\mathcal{A}_1, L_1)$ is [properly] *more specific than* an argument $(\mathcal{A}_2, L_2)$ if, for each activation set H_1 for $(\mathcal{A}_1, L_1)$, there is an activation set $H_2 \subseteq \mathfrak{T}_{H_1 \cup \Pi^G}$ for $(\mathcal{A}_2, L_2)$ [but not vice versa].

Note that this notion of specificity is preliminary, and that the notion of an activation set for an argument has not been properly defined yet.

4.4.2 A first approximation of Activation Sets

In a first approximation, we may now take the labels of all leaves of all resulting trees as the activation set for the original derivation.

The motivation for this notion of an activation set is that the conjunction of its literals is a weakest precondition for all defeasible parts of the concrete original derivation. If such a logically weakest precondition satisfies the specificity notion of § 4.3.2 as an activation set for an argument $(\mathcal{A}_1, L_1)$ w.r.t. a second argument $(\mathcal{A}_2, L_2)$, then any other precondition for all defeasible parts of the given and-tree will satisfy this notion w.r.t. $(\mathcal{A}_2, L_2)$ a fortiori.[7]

4.4.3 Growth of the Defeasible Parts toward the Leaves

Note that in the set of trees resulting from the procedure described in § 4.4.1, there may well have remained instances of rules from Π^{G} connecting a defeasible root application with the defeasible applications right at the leaves. Thus — to cover the whole defeasible part of the derivation in our abstraction — we have to consider the set $\mathcal{A}\cup\Pi^{\mathrm{G}}$ instead of just the set $\mathcal{A}$.

More precisely, we have to include all proper rules (i.e. those with non-empty conditions) from Π^{G}, and may also include the literals in Π^{G} because they cannot do any harm.[8]

As a consequence, in the modeling via our abstraction $\mathcal{A}$, we cannot prevent the isolated defeasible sub-trees resulting from the procedure described in § 4.4.1 from using the rules from Π^{G} to grow toward the root and toward the leaves again.[9] Only the growth toward the leaves, however, can affect our activation sets (which are still taken to be the labels of all leaves of all resulting trees) and thereby our notion of specificity. Indeed, a growth toward the root can add to the conjunction of the given leaves only its super-conjunctions, which are irrelevant because of our focus on weakest preconditions (explained in § 4.4.2)

Let us have a closer look at the effects of such a growth toward the leaves in the most simple case. In addition to a given activation set $\{\mathsf{Q}(\mathsf{a})\}$, in the presence of a general rule

$$\mathsf{Q}(x) \Leftarrow \mathsf{P}_0(x) \wedge \cdots \wedge \mathsf{P}_{n-1}(x)$$

from Π^{G}, we will also have to consider the activation set $\{\ \mathsf{P}_i(\mathsf{a}) \mid i \in \{0, \ldots, n-1\}\ \}$.

This has two effects, which we will discuss in §§ 4.4.4 and 4.4.5.

[7] Note that a further dissection of the isolated defeasible parts would not in general result in activation sets that can be inferred from the strict rules in Π. Where this inference is possible, however, a further dissection leads to the special notion of activation sets given in Definition 8.7 of § 8.3.1.

[8] The need to include all proper rules and to exclude the literals from Π^{F} provides a motivation for simply defining Π^{G} to contain exactly the proper rules of Π, such as found in [STOLZENBURG &AL., 2003].

[9] Of course, our abstraction admits even different defeasible parts of a different and-tree that derives the same literal in focus from the same set $\mathcal{A}$ of instances of defeasible rules, i.e. different derivations of L from $\mathcal{A}\cup\Pi$ for the same argument $(\mathcal{A}, L)$. The admission of these multiple derivations is actually intended in our model-theoretic treatment. The only effect on our current discussion, however, is that we would have to treat several trees disjunctively, which actually makes no difference for the ideas we are currently trying to point out.

[difficult to understand, therefore omit it]

The problem now is that the statement

$$\mathsf{Q}(\mathsf{a}) \not\models \mathsf{P}_0(\mathsf{a}) \wedge \cdots \wedge \mathsf{P}_{n-1}(\mathsf{a}),$$

— which is required to justify this preference — is not explicitly given by the specification $(\Pi^{\mathrm{F}}, \Pi^{\mathrm{G}}, \Delta)$.

Nevertheless — if we do not just want to see it as a matter-of-fact property of notions of specificity in the style of POOLE — we could justify the preference of the "more concise" by imposing the following best practice on positive-conditional specification:

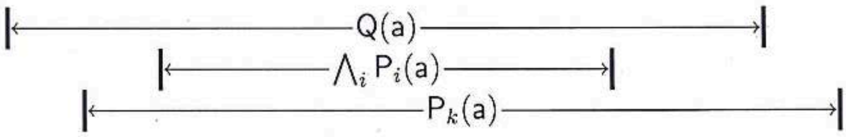
If we write an implication in form of a rule

$$\mathsf{Q}(x) \Leftarrow \mathsf{P}_0(x) \wedge \cdots \wedge \mathsf{P}_{n-1}(x)$$

into a positive-conditional specification Π of strict (i.e. non-defeasible) knowledge, and if we do not intend that the implication is proper in the sense that its converse does not hold in general, then we ought to specify the full equivalence by adding the rules $\mathsf{P}_i(x) \Leftarrow \mathsf{Q}(x)$ $(i \in \{0, \ldots, n-1\})$ to the specification.[12]

Under this best practice of specification, if we find such a rule without the specification of its full equivalence, then it is not intended to exclude models where Q holds for some object a, but not all of the P_i do. This means that if we find such a rule in the strict and general part Π^{G} of a specification, then it is reasonable to assume that the implication is proper w.r.t. the intuition captured in the defeasible rules in Δ.

As a consequence, it makes sense to consider a defeasible argument based on $\{\ \mathsf{P}_i(\mathsf{a}) \mid i \in \{0, \ldots, n-1\}\ \}$ to be properly more specific than an argument that can get along with $\mathsf{Q}(\mathsf{a})$.



Remark 4.2 (Justification for Preference of the "More Concise" Not Valid for Defeasible Rules) ?

Note that our justification for the preference of the "more concise" does not apply, however, if $\mathsf{Q}(x) \Leftarrow \mathsf{P}_0(x) \wedge \cdots \wedge \mathsf{P}_{n-1}(x)$ is a *defeasible* rule instead of a strict one, because we then have the following three problems when trying to justify preference of the "more concise":

- The implication given by the rule is not generally intended (otherwise the rule should be a strict one).
- Moreover, we cannot easily describe the actual instances to which the default rule is meant to apply (otherwise this more concrete description of the defeasible rule should be stated as strict rules).
- The direct treatment of a defeasible equivalence neither has to be appropriate as a default rule in the given situation, nor do we have any means to express a defeasible equivalence in the current setting.

Accordingly, there is, for instance, no clear reason to prefer the first argument of Example 3.3 in § 3 to the second one. This will be discussed in more detail in Example 6.20.

5 Requirements Specification of Specificity in Positive-Conditional Specification

With implicit reference to a defeasible specification $(\Pi^{\mathrm{F}}, \Pi^{\mathrm{G}}, \Delta)$ (cf. § 2.3), let us designate POOLE's relation of being more (or equivalently) specific by "$\lesssim_{\mathrm{P1}}$". Here, "P1" stands for "POOLE's original version".

The standard usage of the symbol "$\lesssim$" is to denote a *quasi-ordering* (cf. § 2.5). Instead of the symbol "$\lesssim$", however, [POOLE, 1985] uses the symbol "$\leq$". The standard usage of the symbol "$\leq$" is to denote a *reflexive ordering* (cf. § 2.5). We cannot conclude from this, however, that POOLE intended the additional property of anti-symmetry; indeed, we find a concrete example specification in [POOLE, 1985] where the lack of anti-symmetry of $\lesssim_{\mathrm{P1}}$ is made explicit.[14]

The possible lack of anti-symmetry of quasi-orderings — i.e. that different arguments may have an equivalent specificity — cannot be a problem because any quasi-ordering $\lesssim_N$ immediately provides us with its equivalence $\approx_N$, its ordering $<_N$, and its reflexive ordering $\leq_N$ (cf. Corollary 2.9 of § 2.5).

By contrast to the non-intended anti-symmetry, *transitivity* is obviously a *conditio sine qua non* for any useful notion of specificity. Indeed, if we have to make a quick choice among the three mutually exclusive actions Propose, Kiss, Smile, and if we already have an argument $(\mathcal{A}_2, \mathsf{Kiss})$ that is more specific than another argument $(\mathcal{A}_3, \mathsf{Smile})$, and if we come up with yet another argument $(\mathcal{A}_1, \mathsf{Propose})$ that is even more specific than $(\mathcal{A}_2, \mathsf{Kiss})$, then, by all means, $(\mathcal{A}_1, \mathsf{Propose})$ should be more specific than the argument $(\mathcal{A}_3, \mathsf{Smile})$ as well. It is obvious that a notion of specificity without transitivity could hardly be helpful in practice.

A further *conditio sine qua non* for any useful notion of specificity is that the conjunctive combination of respectively more specific arguments results in a more specific argument. Indeed, if a square is more specific than a rectangle and a circle is more specific than an ellipse, then a square inscribed into a circle should be more specific than a rectangle inscribed into an ellipse. This property is called monotonicity of conjunction, which we will \emph discuss in § 7.1. Already in [POOLE, 1985], we find an example[15] where $\lesssim_{\mathrm{P1}}$ violates this monotonicity property of the conjunction, which is described there as "seemingly unintuitive".[16]

Further intricacies of computing POOLE's specificity in concrete examples are described in [STOLZENBURG &AL., 2003],[17] which will make it hard to implement $\lesssim_{\mathrm{P1}}$ or its minor corrections as efficiently as required in the practice of answer computation and SLD-resolution w.r.t. positive-conditional specifications.

---

[14] Here we refer to the last three sentences of § 3.2 on Page 145 of [POOLE, 1985].

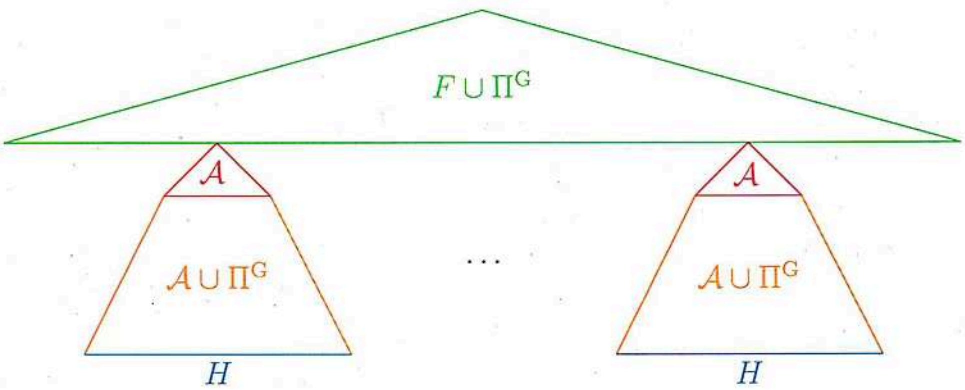
[15] Here we refer to Example 6 of [POOLE, 1985, § 3.5, p.146], which we present here as our Example 7.1 in § 7.1.

[16] See our Example 7.1 in § 7.1 and the references there.

[17] Here we refer to § 3.2ff. of [STOLZENBURG &AL., 2003], where it is demonstrated that, for deciding POOLE's specificity relation (actually $\lesssim_{\mathrm{P2}}$ instead of $\lesssim_{\mathrm{P1}}$, but this does not make any difference here) for two input arguments, we sometimes have to consider even those defeasible rules which are not part of any of these arguments. See also our Example 7.4 in § 7.2.

Activation sets that are not simplified differ from simplified ones by the admission of facts from Π^{F} (in addition to the general rules Π^{G}) after the defeasible part of the derivation is completed.[18]

Our introduction of activation sets that are not simplified is a conceptually important correction of POOLE's approach: It must be admitted to use the facts besides the general rules in a purely strict derivation that is based on literals resulting from completed defeasible arguments, simply because the defeasible parts of a derivation (as isolated in § 4.4.1) should not get more specific by the later use of additional facts that do not provide input to the defeasible parts.[19] Note that the difference between simplified and non-simplified activation sets typically occurs in real applications, but — except Example 7.5 in § 7.2 — not in our toy examples of § 7, which mainly exemplify the differences in phase 1.



	$H \subseteq \ldots$	$F = \ldots$
POOLE's Approach (P1, P2, P3)	$\mathfrak{T}_{\Pi^{\mathrm{F}} \cup \Pi^{\mathrm{G}} \cup \Delta}$	H [20]
Our Approach (CP1, CP2)	$\mathfrak{T}_{\Pi^{\mathrm{F}} \cup \Pi^{\mathrm{G}}}$	Π^{F} [21]

Figure 1: And-Tree with Phases 1, 2, 3.[22]

[18]This can be seen in Example 7.5 of § 7, and in Example 8.1 of § 8.2.2. See also the variable F in Figure 1.

[19]We do not further discuss this obviously appropriate correction here and leave the construction of examples that make the conceptual necessity of this correction intuitively clear as an exercise. Hint: Have a look at the proof of Theorem 6.16 in § 6.5. Then present two different sets of strict rules with equal derivability, where only one needs the facts in phase 3 and where the additional specificity gained by these facts violates the intuition.

[20]Look at Note 35 of Example 7.4 in § 7.2 to see that it may really matter for the definition of P1, P2, P3 that we do *not* have $H \subseteq \mathfrak{T}_{\Pi^{\mathrm{F}} \cup \Pi^{\mathrm{G}}}$ in general in POOLE's approach.

[21]Although we do *not* have $H \subseteq \Pi^{\mathrm{F}}$ in general in our approach, the replacement of Π^{F} with H in this table would result in fewer derivable roots for our approach, simply because we always have $\mathfrak{T}_{H \cup \Pi^{\mathrm{G}}} \subseteq \mathfrak{T}_{\Pi^{\mathrm{F}} \cup \Pi^{\mathrm{G}}}$ in our approach.

[22]From leaves to the root: phase 1 (H), phase 2 (sub-trees of the defeasible parts of a derivation, with explicit defeasible root steps), phase 3 (root sub-tree).

We have $\mathfrak{T}_{\Pi_{6.5}} = \{\mathsf{thirst}, \mathsf{drink}\}$, $\mathfrak{T}_{\Pi_{6.5} \cup \Delta_{6.5}} = \{\mathsf{beer}\} \cup \mathfrak{T}_{\Pi_{6.5}}$.

We have $(\mathcal{A}_2, \mathsf{beer}) \lesssim_{\mathrm{P2}} (\emptyset, \mathsf{drink})$ because for every $H \subseteq \mathfrak{T}_{\Pi_{6.5} \cup \Delta_{6.5}}$ that is a simplified activation set for $(\mathcal{A}_2, \mathsf{beer})$, but not a simplified activation set for $(\emptyset, \mathsf{beer})$, we have $H = \{\mathsf{thirst}\}$, which is a simplified activation set also for $(\emptyset, \mathsf{drink})$.

We have $(\emptyset, \mathsf{drink}) \lesssim_{\mathrm{P2}} (\mathcal{A}_2, \mathsf{beer})$ because there cannot be a simplified activation set for $(\emptyset, \mathsf{drink})$ that is not a simplified activation set for $(\emptyset, \mathsf{drink})$.

All in all, we get[23] $(\mathcal{A}_2, \mathsf{beer}) \approx_{\mathrm{P2}} (\emptyset, \mathsf{drink})$, although $(\emptyset, \mathsf{drink}) <_{\mathrm{P3}} (\mathcal{A}_2, \mathsf{beer})$ should be given according to intuition, because an argument that does not require any defeasible rules should be strictly preferred to an argument that does: If beer produces a conflict with our drinking habits, there is no reason to prefer it to any other drink.

[is not yet introduced, [illegible] be omit P3 here]

To overcome this minor flaw, which consists in the inconvenience of not in general preferring a non-defeasible argument to a comparable defeasible one, we finally add an implication as an additional requirement in Definition 6.6. This implication guarantees that no argument that requires defeasible rules can be more or equivalently specific than an argument that does not require any defeasible rules at all.

Definition 6.6 ($\lesssim_{\mathrm{P3}}$: **Rather Unflawed Version of** DAVID POOLE**'s Specificity**)
$(\mathcal{A}_1, L_1) \lesssim_{\mathrm{P3}} (\mathcal{A}_2, L_2)$ if $(\mathcal{A}_1, L_1)$ and $(\mathcal{A}_2, L_2)$ are arguments, $L_2 \in \mathfrak{T}_\Pi$ implies $L_1 \in \mathfrak{T}_\Pi$, and if, for every $H \subseteq \mathfrak{T}_{\Pi \cup \Delta}$ that is a [minimal][24] simplified activation set for $(\mathcal{A}_1, L_1)$ but not a simplified activation set for $(\emptyset, L_1)$, H is also a simplified activation set for $(\mathcal{A}_2, L_2)$.

Corollary 6.7 *If $(\mathcal{A}_1, L_1)$, $(\mathcal{A}_2, L_2)$ are arguments with $\mathcal{A}_1 \subseteq \mathcal{A}_2$, then any of the following conditions is sufficient for $(\mathcal{A}_1, L_1) \lesssim_{\mathrm{P3}} (\mathcal{A}_2, L_2)$:*

1. $L_1 = L_2$.
2. $L_2 \in \mathfrak{T}_\Pi \Rightarrow L_1 \in \mathfrak{T}_\Pi$ *and* $\{L_1\} \cup \mathcal{A}_2 \cup \Pi^{\mathrm{G}} \vdash \{L_2\}$,
3. $\mathcal{A}_1 = \emptyset$ *(which implies $L_1 \in \mathfrak{T}_\Pi$ by Definition 2.6).*[25]

As every simplified activation set that passes the condition of Definition 6.3 also passes the one of Definitions 6.4 and 6.6, we get the following corollary of these three definitions.

Corollary 6.8 $\lesssim_{\mathrm{P3}} \ \subseteq \ \lesssim_{\mathrm{P2}} \ \subseteq \ \lesssim_{\mathrm{P1}}$.

By Corollaries 6.7 and 6.8, $\lesssim_{\mathrm{P1}}$, $\lesssim_{\mathrm{P2}}$, $\lesssim_{\mathrm{P3}}$ are reflexive relations on arguments, but — as we will show in Example 6.9 and state in Theorem 6.11 — not quasi-orderings in general.

[23] Note that by Corollary 6.8, we will get $(\mathcal{A}_2, \mathsf{beer}) \approx_{\mathrm{P1}} (\emptyset, \mathsf{drink})$ as well. Finally note that this problem does not occur in the similar Example 3.1 of § 3.

[difficult to understand and not neces]

[24] Note that the omission of the optional restriction to *minimal* simplified activation sets for $(\mathcal{A}_1, L_1)$ in Definition 6.6 has no effect on the extension of the defined notion, simply because the additional non-minimal simplified activation sets $(\mathcal{A}_1, L_1)$ will then be simplified activation sets for $(\mathcal{A}_2, L_2)$ *a fortiori*.

[25] Exercise: Find a counterexample, however, for the conjecture that $L_1 \in \mathfrak{T}_\Pi$ implies $(\mathcal{A}, L_1) \lesssim_{\mathrm{P3}} (\mathcal{A}, L_2)$.

Proof of Lemma 6.10

Looking at Example 6.9, we see that only the quasi-ordering properties in the last two lines of Lemma 6.10 are non-trivial. We have

$$\mathfrak{T}_{\Pi_{6.9}} = \{\mathsf{Bold},\ \mathsf{HAkimbo(Jo)},\ \mathsf{Smiles(Jo)},\ \mathsf{Sexy(Jo)}\},$$
$$\mathfrak{T}_{\Pi_{6.9}\cup\Delta_{6.9}} = \{\mathsf{Promising(Jo)},\ \mathsf{Propose},\ \mathsf{Kiss},\ \mathsf{Smile}\} \cup \mathfrak{T}_{\Pi_{6.9}}.$$

Thus, regarding the arguments $(\mathcal{A}_1, \mathsf{Propose})$, $(\mathcal{A}_2, \mathsf{Kiss})$, $(\mathcal{A}_3, \mathsf{Smile})$, the implication added in Definition 6.6 as compared to Definitions 6.3 and 6.4 is always satisfied, simply because its condition is always false.

$(\mathcal{A}_3, \mathsf{Smile}) \not\gtrsim_{\mathrm{P1}} (\mathcal{A}_1, \mathsf{Propose}) \lesssim_{\mathrm{P3}} (\mathcal{A}_2, \mathsf{Kiss})$: The minimal simplified activation sets for $(\mathcal{A}_1, \mathsf{Propose})$ that are subsets of $\mathfrak{T}_{\Pi_{6.9}\cup\Delta_{6.9}}$ and no simplified activation sets for $(\emptyset, \mathsf{Propose})$ (or, without any difference, no simplified activation sets for $(\mathcal{A}_3, \mathsf{Propose})$) are $\{\mathsf{Bold}, \mathsf{HAkimbo(Jo)}, \mathsf{Smiles(Jo)}, \mathsf{Sexy(Jo)}\}$ and $\{\mathsf{Bold}, \mathsf{Promising(Jo)}\}$, which are simplified activation sets for $(\mathcal{A}_2, \mathsf{Kiss})$ — but $\{\mathsf{Bold}, \mathsf{Promising(Jo)}\}$ is no simplified activation set for $(\mathcal{A}_3, \mathsf{Smile})$.

$(\mathcal{A}_1, \mathsf{Propose}) \not\gtrsim_{\mathrm{P1}} (\mathcal{A}_2, \mathsf{Kiss}) \lesssim_{\mathrm{P3}} (\mathcal{A}_3, \mathsf{Smile})$: The only simplified activation set for $(\mathcal{A}_2, \mathsf{Kiss})$ that is a subset of $\mathfrak{T}_{\Pi_{6.9}\cup\Delta_{6.9}}$ and no simplified activation set for $(\emptyset, \mathsf{Kiss})$ (such as $\{\mathsf{Promising(Jo)}\}$) (or, without any difference, no simplified activation sets for $(\mathcal{A}_1, \mathsf{Kiss})$) is $\{\mathsf{Bold}, \mathsf{Smiles(Jo)}, \mathsf{Sexy(Jo)}\}$, which is a simplified activation set for $(\mathcal{A}_3, \mathsf{Smile})$, but not for $(\mathcal{A}_1, \mathsf{Propose})$.

$(\mathcal{A}_2, \mathsf{Kiss}) \not\gtrsim_{\mathrm{P1}} (\mathcal{A}_3, \mathsf{Smile})$: The only minimal simplified activation set for $(\mathcal{A}_3, \mathsf{Smile})$ that is a subset of $\mathfrak{T}_{\Pi_{6.9}\cup\Delta_{6.9}}$ and no simplified activation set for $(\mathcal{A}_2, \mathsf{Smile})$ is $\{\mathsf{Sexy(Jo)}\}$, which is not a simplified activation set for $(\mathcal{A}_2, \mathsf{Kiss})$.

Q.e.d. (Lemma 6.10)

6.3 Main Negative Result: Not Transitive!

The relations stated in Lemma 6.10 hold not only for the given indices, but — by Corollary 6.8 — actually for all of P1, P2, P3; and so we immediately get:

Theorem 6.11
There is a specification $(\Pi^{\mathrm{F}}_{6.9}, \Pi^{\mathrm{G}}_{6.9}, \Delta_{6.9})$, such that $\Pi^{\mathrm{F}}_{6.9} \cup \Pi^{\mathrm{G}}_{6.9} \cup \Delta_{6.9}$ is non-contradictory, but none of $\lesssim_{\mathrm{P1}}$, $\lesssim_{\mathrm{P2}}$, $\lesssim_{\mathrm{P3}}$, $<_{\mathrm{P1}}$, $<_{\mathrm{P2}}$, $<_{\mathrm{P3}}$ is transitive. Moreover, the counterexamples to the transitivity of all these relations can be restricted to minimal arguments.

As a consequence of Theorem 6.11, the respective relations in [POOLE, 1985], [STOLZENBURG &AL., 2003], and [SIMARI & LOUI, 1992] are not transitive. This means that these relations are not quasi-orderings, let alone reflexive orderings.

This consequence is immediate for the relation $\geq$ at the bottom of the left column on Page 145 of [POOLE, 1985]. Moreover, note that the consequence does not depend on the contentious question on whether our interpretation of the negation symbol $\neg$ essentially differs from its interpretation in [POOLE, 1985]. Indeed, our counterexample to transitivity occurs in the negation-free definite-rule fragment of POOLE's original language.

Corollary 6.13 *If $(\mathcal{A}_1, L_1)$, $(\mathcal{A}_2, L_2)$ are arguments with $\mathcal{A}_1 \subseteq \mathcal{A}_2$, then any of the following conditions is sufficient for $(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP1}} (\mathcal{A}_2, L_2)$:*

1. $L_1 = L_2$.
2. $L_2 \in \mathfrak{T}_\Pi \Rightarrow L_1 \in \mathfrak{T}_\Pi$ *and* $\{L_1\} \cup \Pi \vdash \{L_2\}$.[29]
3. $L_1 \in \mathfrak{T}_\Pi$ *(which is implied by $\mathcal{A}_1 = \emptyset$ by Definition 2.6).*

The crucial change in Definition 6.12 as compared to Definition 6.6 is *not* the technically required emphasis it puts on the case "$L_1 \in \mathfrak{T}_\Pi$"[, which will be discussed in Remark 6.18 of § 6.6.] The crucial changes actually are [may be omitted, too many forward references]

(A) the replacement of "$H \subseteq \mathfrak{T}_{\Pi\cup\Delta}$" with "$H \subseteq \mathfrak{T}_\Pi$" (as explained already in phase 1 of § 6.1), and the thereby enabled

(B) omission of the previously technically required,[30] but unintuitive negative condition on derivability (of the form "but not a simplified activation set for $(\emptyset, L_1)$").

An additional minor change, which we have already discussed in § 6.1, is the one from simplified activation sets to (non-simplified) activation sets.

Theorem 6.14 $\precsim_{\mathrm{CP1}}$ *is a quasi-ordering on arguments.*

Proof of Theorem 6.14

$\precsim_{\mathrm{CP1}}$ is a reflexive relation on arguments because of Corollary 6.13.

To show transitivity, let us assume

$$(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP1}} (\mathcal{A}_2, L_2) \text{ and } (\mathcal{A}_2, L_2) \precsim_{\mathrm{CP1}} (\mathcal{A}_3, L_3).$$

According to Definition 6.12, because of $(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP1}} (\mathcal{A}_2, L_2)$, we have $L_1 \in \mathfrak{T}_\Pi$ — and then immediately the desired $(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP1}} (\mathcal{A}_3, L_3)$ — or we have $L_2 \notin \mathfrak{T}_\Pi$ and every $H \subseteq \mathfrak{T}_\Pi$ that is an activation set for $(\mathcal{A}_1, L_1)$ is also an activation set for $(\mathcal{A}_2, L_2)$. The latter case excludes the first option in Definition 6.12 as a justification for $(\mathcal{A}_2, L_2) \precsim_{\mathrm{CP1}} (\mathcal{A}_3, L_3)$, and thus we have $L_3 \notin \mathfrak{T}_\Pi$ and every $H \subseteq \mathfrak{T}_\Pi$ that is an activation set for $(\mathcal{A}_2, L_2)$ is also an activation set for $(\mathcal{A}_3, L_3)$. All in all, we get that every $H \subseteq \mathfrak{T}_\Pi$ that is an activation set for $(\mathcal{A}_1, L_1)$ is also an activation set for $(\mathcal{A}_3, L_3)$. Thus, we get the desired $(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP1}} (\mathcal{A}_3, L_3)$ also in this case. **Q.e.d. (Theorem 6.14)**

Obviously, an argument is ranked by $\precsim_{\mathrm{CP1}}$ firstly on whether its literal is in $\mathfrak{T}_\Pi$, and, if not, secondly on the set of its activation sets, which is ~~an element of the power set~~ a subset of the power set of $\mathfrak{T}_\Pi$. So we get:

Corollary 6.15 *If $\mathfrak{T}_\Pi$ is finite, then $<_{\mathrm{CP1}}$ is well-founded.*

---

[29] Note that, in general — contrary to Corollary 6.7(2) — $\mathcal{A}_2$ must not participate in the derivation of L_2 from L_1, say in the form that there is a set of literals $\mathfrak{L}$ with $\{L_1\} \cup \mathcal{A}_2 \cup \Pi^{\mathrm{G}} \vdash \mathfrak{L}$ and $\mathfrak{L} \cup \Pi \vdash \{L_2\}$, because rules from Π^{F} may have participated in the derivation of L_1 from an activation set. The source of this difference between P3 and CP1 is the replacement of simplified activation sets in Definition 6.6 with (non-simplified) activation sets in Definition 6.12.

[30] See the discussion in Example 6.21 in § 6.6 on why this condition is technically required for P1, P2, and P3.

Let us now provide an *argumentum ad absurdum* for the assumption that H is a simplified activation set also for $(\emptyset, L_1)$: Then we would have $L_1 \in \mathfrak{T}_{H \cup \Pi^{\mathrm{G}}}$, and because of $H \subseteq \mathfrak{T}_\Pi$ and $\Pi^{\mathrm{G}} \subseteq \Pi$ we get $L_1 \in \mathfrak{T}_{\mathfrak{T}_\Pi \cup \Pi} = \mathfrak{T}_\Pi$ — a contradiction to our current case of $L_1, L_2 \not\in \mathfrak{T}_\Pi$.

All in all, by our initial assumption, H must now be a simplified activation set for $(\mathcal{A}_2, L_2)$ and, *a fortiori* by Corollary 6.2, an activation set for $(\mathcal{A}_2, L_2)$, as was to be shown for our only remaining sub-claim. **Q.e.d. (Theorem 6.16)**

6.6 Checking Up the Previous Examples

With the help of Theorem 6.16, we can now analyze the examples of § 3, and also check how our relation CP1 behaves in case of our counterexample to transitivity. Note that condition 4 of Theorem 6.16 is satisfied for all of these examples.

Example 6.17 *(continuing Example 3.1 of § 3)*
We have $(\mathcal{A}_2, \mathsf{flies(edna)}) \not\lesssim_{\mathrm{CP1}} (\emptyset, \neg\mathsf{flies(edna)})$
because $\mathsf{flies(edna)} \not\in \mathfrak{T}_{\Pi_{3.1}}$ and $\neg\mathsf{flies(edna)} \in \mathfrak{T}_{\Pi_{3.1}}$.
We have $(\emptyset, \neg\mathsf{flies(edna)}) \lesssim_{\mathrm{P3}} (\mathcal{A}_2, \mathsf{flies(edna)})$ by Corollary 6.7(3).
All in all, by Theorem 6.16, we get $(\emptyset, \neg\mathsf{flies(edna)}) <_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{flies(edna)})$
and $(\emptyset, \neg\mathsf{flies(edna)}) <_{\mathrm{P3}} (\mathcal{A}_2, \mathsf{flies(edna)})$.

Remark 6.18 One may ask why we did not define an additional quasi-ordering, say $\lesssim_{\mathrm{CP0}}$, simply by replacing the two conditions of Definition 6.12 with the single condition

"$L_2 \in \mathfrak{T}_\Pi$ implies $L_1 \in \mathfrak{T}_\Pi$, and every $H \subseteq \mathfrak{T}_\Pi$ that is an [minimal] activation set for $(\mathcal{A}_1, L_1)$ is also an activation set for $(\mathcal{A}_2, L_2)$."

This would be more in the style of Definition 6.6 for $\lesssim_{\mathrm{P3}}$, and would also avoid the singular behavior of the first alternative condition of Definition 6.12, and so offer continuity advantages.[31] Moreover, for $\lesssim_{\mathrm{CP0}}$ instead of $\lesssim_{\mathrm{CP1}}$, items 1 and 2 (but not item 3) of Corollary 6.13 still hold, as well as Theorem 6.14 and its Corollary 6.15. Furthermore, we get $\lesssim_{\mathrm{CP0}} \subseteq \lesssim_{\mathrm{CP1}}$. It is fatal for $\lesssim_{\mathrm{CP0}}$, however, that this subset relation may be proper. For instance, $\lesssim_{\mathrm{CP0}}$ does not in general satisfy Theorem 6.16. Even worse, $\lesssim_{\mathrm{CP0}}$ does not show the proper behavior of $\lesssim_{\mathrm{CP1}}$ in Example 3.1 of § 3, as discussed in Example 6.17 of § 6.6: We get $(\emptyset, \neg\mathsf{flies(edna)}) \,\Delta_{\mathrm{CP0}}\, (\mathcal{A}_2, \mathsf{flies(edna)})$ instead of
$(\emptyset, \neg\mathsf{flies(edna)}) <_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{flies(edna)})$.

This can be seen by considering the activation set $\emptyset$ for $(\emptyset, \neg\mathsf{flies(edna)})$, which is not an activation set for $(\mathcal{A}_2, \mathsf{flies(edna)})$.

Such a behavior is obviously unacceptable in practice, and so we do not think that it makes sense to consider $\lesssim_{\mathrm{CP0}}$ any further. [may be omitted]

Example 6.19 *(continuing Example 3.2 of § 3)*
We have $(\mathcal{A}_2, \mathsf{flies(edna)}) \not\lesssim_{\mathrm{CP1}} (\mathcal{A}_1, \neg\mathsf{flies(edna)})$ because $\mathsf{flies(edna)} \not\in \mathfrak{T}_{\Pi_{3.2}}$ and because $\{\mathsf{bird(edna)}\} \subseteq \mathfrak{T}_{\Pi_{3.2}}$ is an activation set for $(\mathcal{A}_2, \mathsf{flies(edna)})$, but not for $(\mathcal{A}_1, \neg\mathsf{flies(edna)})$.

We have $(\mathcal{A}_1, \neg\mathsf{flies(edna)}) \lesssim_{\mathrm{P3}} (\mathcal{A}_2, \mathsf{flies(edna)})$, because $\mathsf{flies(edna)} \not\in \mathfrak{T}_{\Pi_{3.2}}$ and because, if $H \subseteq \mathfrak{T}_{\Pi_{3.2} \cup \Delta_{3.2}}$ is a simplified activation set for $(\mathcal{A}_1, \neg\mathsf{flies(edna)})$, but not for

[31] Cf. the discussion of such a continuity advantage in § 7.1 for the monotonicity w.r.t. conjunction.

Example 6.22 *(continuing Example 6.9 of § 6.2)*
The following holds for our specification of Example 6.9 by Lemma 6.10 and Corollary 6.8:

$$(\mathcal{A}_1, \mathsf{Propose}) <_{\mathrm{P3}} (\mathcal{A}_2, \mathsf{Kiss}) <_{\mathrm{P3}} (\mathcal{A}_3, \mathsf{Smile}) \not\gtrsim_{\mathrm{P3}} (\mathcal{A}_1, \mathsf{Propose}).$$

For our corrected relation CP1 we have:

$$(\mathcal{A}_1, \mathsf{Propose}) <_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{Kiss}) <_{\mathrm{CP1}} (\mathcal{A}_3, \mathsf{Smile}) >_{\mathrm{CP1}} (\mathcal{A}_1, \mathsf{Propose})$$

simply because the trouble-making set $\{\,\mathsf{Bold},\ \mathsf{Promising(Jo)}\,\}$ is not to be considered here. Indeed, this set is not a subset of $\mathfrak{T}_{\Pi_{6.9}}$. The checking of the details is left to the reader. Note that, because of Lemma 6.10, Theorem 6.16, Theorem 6.14, and Corollary 2.9, all that is actually left to show is

$$(\mathcal{A}_1, \mathsf{Propose}) \not\gtrsim_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{Kiss}) \not\gtrsim_{\mathrm{CP1}} (\mathcal{A}_3, \mathsf{Smile}).$$

7 Putting Specificity to Test w.r.t. Human Intuition

Before we will go on with further conceptual material and efficiency considerations in § 8, let us put our two main notions of specificity — as formalized in the two binary relations $\lesssim_{\mathrm{P3}}$ and $\lesssim_{\mathrm{CP1}}$ — to test w.r.t. our changed phase 1 of § 6.1 in a series of further examples.

Note that we can freely draw the consequence $\lesssim_{\mathrm{P3}} \subseteq \lesssim_{\mathrm{CP1}}$ of Theorem 6.16 because at least one[33] of its conditions is satisfied in all the following examples except Example 7.5, which is the only example in § 7 with an activation set that actually is not a simplified one.

Besides freely applying Theorem 6.16 — to enable the reader to make his own selection of interesting examples — we are pretty explicit in all of the following examples.

7.1 Monotonicity of the Specificity Relations w.r.t. Conjunction

Monotonicity w.r.t. conjunction is the following property for a binary relation R on arguments: In case of $(\mathcal{A}_1^i, L_1^i)\ R\ (\mathcal{A}_2^i, L_2^i)$ for $i \in \{1,2\}$,
we always have $(\mathcal{A}_1^1 \cup \mathcal{A}_1^2, L_1')\ R\ (\mathcal{A}_2^1 \cup \mathcal{A}_2^2, L_2')$
for fresh constant literals L_j' with rules $L_j' \Leftarrow L_j^1 \wedge L_j^2$ added to the general rules Π^{G} ($j \in \{1,2\}$). In this case, we will call $(\mathcal{A}_j^1 \cup \mathcal{A}_j^2, L_1')$ the *conjunction* of the arguments $(\mathcal{A}_j^1, L_j^1)$ and $(\mathcal{A}_j^2, L_j^2)$.

This property is obviously given for $\lesssim_{\mathrm{CP1}}$ in case of $L_1^1, L_1^2 \in \mathfrak{T}_\Pi$ (which implies $L_1' \in \mathfrak{T}_\Pi$) and also in case of $L_1^1, L_1^2 \notin \mathfrak{T}_\Pi$ (where we get $L_2^1, L_2^2, L_1', L_2' \notin \mathfrak{T}_\Pi$ and just take the union of the two activation sets). Note that the latter case — where both arguments are defeasible — is certainly the most important one.

For the remaining borderline case of $L_1^i \notin \mathfrak{T}_\Pi \ni L_1^{3-i}$ (for some $i \in \{1,2\}$), however, monotonicity cannot be expected in general for $\lesssim_{\mathrm{CP1}}$, simply because then we get $L_1' \notin \mathfrak{T}_\Pi$, but do not necessarily have any activation set for L_2^{3-i}. This non-monotonicity, however, is part and parcel of our decision to prefer arguments whose literals are elements of $\mathfrak{T}_\Pi$, as expressed in item 1 of Definition 6.12 of § 6.4. As explained in Remark 6.18 of § 6.6, there does not seem to be an alternative to this technically required preference.

For $\lesssim_{\mathrm{P1}}$, however, monotonicity is not even given for the case we just realized to be the most important one. This was already noted in [POOLE, 1985], using the following example:

[33] Condition 4 of Theorem 6.16 is satisfied for Examples 3.2, 3.3, 3.4, and 7.7. Condition 3 (but not condition 4) is satisfied for Examples 7.1, 7.2, 7.3, 7.4 and 7.6.

[may be omitted]

the same reason.

We have $(\mathcal{A}_1, \mathsf{g}_1) <_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{g}_2)$, which is intuitively correct because the conjunction of a more specific and an equivalently specific argument, respectively, should be more specific. Indeed, from the isomorphic sub-specifications in Examples 3.2 and 3.3, we know that $(\mathcal{A}_1, \neg\mathsf{c}) <_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{c})$ and $(\mathcal{A}_1, \neg\mathsf{f}) \approx_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{f})$, respectively.

All in all, the relation $\lesssim_{\mathrm{P3}}$ fails in this example again, whereas the quasi-ordering $\lesssim_{\mathrm{CP1}}$ works according to human intuition and satisfies the required monotonicity w.r.t. conjunction of § 5.

7.2 Implementation of the Preference of the "More Precise"

As primary sources of differences in specificity, all previous examples — except Example 3.4 of § 3, continued in Example 6.21 of § 6.6 — illustrate only the effect of chains of implications. According to our motivating discussion of § 4.4.5, we should consider also examples where the primary source of differences in specificity is an essentially required condition that is a super-conjunction of the condition triggering another rule. We will do so in the following examples.

As we have already shown in Example 6.21, both relations $\lesssim_{\mathrm{P3}}$ and $\lesssim_{\mathrm{CP1}}$ produce the intuitive result if the "more precise" super-conjunction is *directly* the condition of a rule. Let us see whether this is also the case if the condition of the rule is *derived* from a super-conjunction.

By removing the second condition literal $\neg\mathsf{f}$ in the strict general rule $\mathsf{g}_1 \Leftarrow \neg\mathsf{c} \wedge \neg\mathsf{f}$ of Example 7.1, we obtain the following example.

Example 7.3 (2nd Variation of Example 7.1)

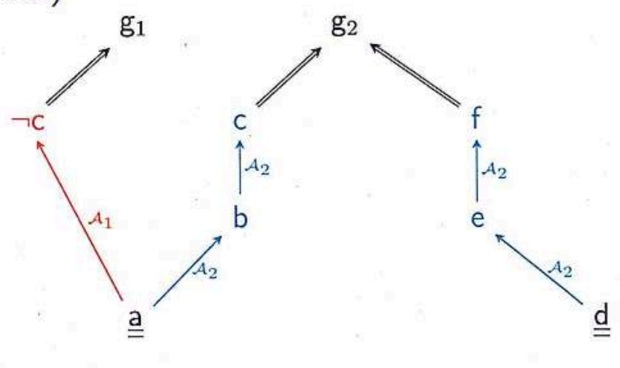
$$\Pi^{\mathrm{F}}_{7.3} := \{\ \mathsf{a},\ \mathsf{d}\ \},$$

$$\Pi^{\mathrm{G}}_{7.3} := \left\{ \begin{array}{l} \mathsf{g}_1 \Leftarrow \neg\mathsf{c}, \\ \mathsf{g}_2 \Leftarrow \mathsf{c} \wedge \mathsf{f} \end{array} \right\},$$

$$\Delta_{7.3} := \mathcal{A}_1 \cup \mathcal{A}_2.$$

$$\mathcal{A}_1 := \{\ \neg\mathsf{c} \leftarrow \mathsf{a}\ \}.$$

$$\mathcal{A}_2 := \left\{ \begin{array}{l} \mathsf{b} \leftarrow \mathsf{a}, \\ \mathsf{c} \leftarrow \mathsf{b}, \\ \mathsf{e} \leftarrow \mathsf{d}, \\ \mathsf{f} \leftarrow \mathsf{e} \end{array} \right\}.$$



Let us compare the specificity of the arguments $(\mathcal{A}_1, \mathsf{g}_1)$ and $(\mathcal{A}_2, \mathsf{g}_2)$.

We have $(\mathcal{A}_1, \mathsf{g}_1) \not\lesssim_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{g}_2)$ because $\{\mathsf{a}\} \subseteq \mathfrak{T}_{\Pi_{7.3}} = \{\mathsf{a}, \mathsf{d}\}$ is an activation set for $(\mathcal{A}_1, \mathsf{g}_1)$, but not for $(\mathcal{A}_2, \mathsf{g}_2)$.

We have $(\mathcal{A}_2, \mathsf{g}_2) \lesssim_{\mathrm{CP1}} (\mathcal{A}_1, \mathsf{g}_1)$ because any activation set for $(\mathcal{A}_2, \mathsf{g}_2)$ that is a subset of $\mathfrak{T}_{\Pi_{7.3}}$ includes a, and so is also an activation set for $(\mathcal{A}_1, \mathsf{g}_1)$.

Considering Theorem 6.16 as well as the the activation set $\{\mathsf{b}, \mathsf{d}\}$ for $(\mathcal{A}_2, \mathsf{g}_2)$,
we get $(\mathcal{A}_1, \mathsf{g}_1)\ \triangle_{\mathrm{P3}}\ (\mathcal{A}_2, \mathsf{g}_2)$,
contrary to $(\mathcal{A}_1, \mathsf{g}_1) >_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{g}_2)$.

Thus, $\lesssim_{\mathrm{CP1}}$ realizes the intuition that the super-conjunction $\mathsf{a} \wedge \mathsf{d}$ — which is essential to derive $\mathsf{c} \wedge \mathsf{f}$ according to $\mathcal{A}_2$ — is more specific than the "less precise" a.

Just like Example 6.20 of § 6.6, this example shows again that $\lesssim_{\mathrm{P3}}$ does not properly

By turning the defeasible rule $\mathsf{f}{\leftarrow}\mathsf{e}$ of Example 7.4 into a strict general rule, we obtain the following example.

Example 7.5 (Variation of Example 7.4)

$\Pi^{\mathrm{F}}_{7.5} \;:=\; \{\, \mathsf{c},\ \mathsf{d},\ \mathsf{e} \,\}$,

$\Pi^{\mathrm{G}}_{7.5} \;:=\; \left\{ \begin{array}{l} \mathsf{x} \Leftarrow \mathsf{a}{\wedge}\mathsf{f}, \\ \mathsf{f} \Leftarrow \mathsf{e} \end{array} \right\}$,

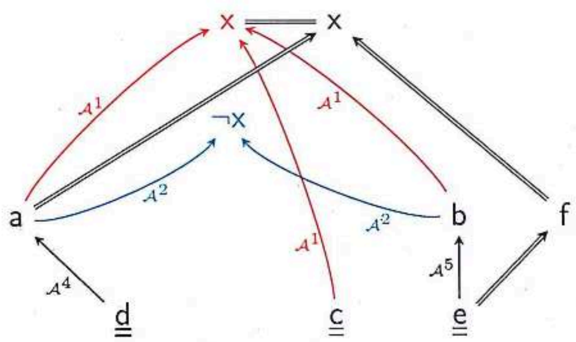
$\Delta_{7.5} \;:=\; \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^4 \cup \mathcal{A}^5$,

$\mathcal{A}^1 \;:=\; \{\, \mathsf{x} \leftarrow \mathsf{a}{\wedge}\mathsf{b}{\wedge}\mathsf{c} \,\}$.

$\mathcal{A}^2 \;:=\; \{\, \neg\mathsf{x} \leftarrow \mathsf{a}{\wedge}\mathsf{b} \,\}$.

$\mathcal{A}^4 \;:=\; \{\, \mathsf{a} \leftarrow \mathsf{d} \,\}$,

$\mathcal{A}^5 \;:=\; \{\, \mathsf{b} \leftarrow \mathsf{e} \,\}$.



Compare the specificity of the arguments $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$, $(\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x})$, $(\mathcal{A}^4, \mathsf{x})$!

Obviously, $\mathsf{x}, \neg\mathsf{x} \notin \mathfrak{T}_{\Pi_{7.5}} = \{\mathsf{c},\mathsf{d},\mathsf{e},\mathsf{f}\}$. Moreover, $\{\mathsf{d}\} \subseteq \mathfrak{T}_{\Pi_{7.5}}$ is an activation set for $(\mathcal{A}^4, \mathsf{x})$ (but not a simplified one!) and, *a fortiori* (by Corollary 6.13(1)), for $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$, but not for $(\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x})$. Furthermore, every activation set $H \subseteq \mathfrak{T}_{\Pi_{7.5}}$ for $(\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x})$ satisfies $\{\mathsf{d},\mathsf{e}\} \subseteq H$, which is an activation set for $(\mathcal{A}^4, \mathsf{x})$ and $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$. Finally, every activation set $H \subseteq \mathfrak{T}_{\Pi_{7.5}}$ for $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$ satisfies $\{\mathsf{d}\} \subseteq H$ which is an activation set for $(\mathcal{A}^4, \mathsf{x})$.

All in all, we have $(\mathcal{A}^4, \mathsf{x}) \approx_{\mathrm{CP1}} (\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x}) >_{\mathrm{CP1}} (\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x})$.

This is intuitively sound because $(\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x})$ is activated only by the more specific $\mathsf{d}{\wedge}\mathsf{e}$, whereas $(\mathcal{A}^4, \mathsf{x})$ is activated also by the "less precise" d.

Moreover, $\mathsf{c}{\wedge}\mathsf{d}{\wedge}\mathsf{e}$ is not essentially required for $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$, and so this argument is tantamount to $(\mathcal{A}^4, \mathsf{x})$. The reason for this remarkable effect is not the lack of minimality of the argument $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$, but our semantical, model-theoretic approach, which simply ignores the fact that the derivation via $\mathcal{A}^1$ requires the more precise activation set. Indeed, we primarily consider consequence, not derivation.

We have $(\mathcal{A}^4, \mathsf{x}) <_{\mathrm{P3}} (\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x}) \vartriangle_{\mathrm{P3}} (\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x}) \vartriangle_{\mathrm{P3}} (\mathcal{A}^4, \mathsf{x})$, however.[36]

This means that $\lesssim_{\mathrm{P3}}$ fails here completely w.r.t. POOLE's intuition, as actually in most non-trivial examples.

[may be omitted]

7.3 Conflict between the "More Concise" and the "More Precise"

By removing the second condition literal $\neg\mathsf{f}$ in the strict general rule $\mathsf{g}_1{\Leftarrow}\neg\mathsf{c}{\wedge}\neg\mathsf{f}$ of Example 7.2, we obtain the following example.

[36]The minimal simplified activation sets for $(\mathcal{A}^4, \mathsf{x})$ that are no simplified activation sets for $(\emptyset, \mathsf{x})$ are $\{\mathsf{d},\mathsf{e}\}$ and $\{\mathsf{d},\mathsf{f}\}$. The minimal simplified activation sets for $(\mathcal{A}^1{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \mathsf{x})$ that are no simplified activation sets for $(\emptyset, \mathsf{x})$ are $\{\mathsf{d},\mathsf{e}\}$, $\{\mathsf{d},\mathsf{f}\}$, $\{\mathsf{a},\mathsf{b},\mathsf{c}\}$, and $\{\mathsf{b},\mathsf{c},\mathsf{d}\}$. The minimal simplified activation sets for $(\mathcal{A}^2{\cup}\mathcal{A}^4{\cup}\mathcal{A}^5, \neg\mathsf{x})$ that are no simplified activation sets for $(\emptyset, \neg\mathsf{x})$ are $\{\mathsf{a},\mathsf{b}\}$, $\{\mathsf{a},\mathsf{e}\}$, $\{\mathsf{b},\mathsf{d}\}$, and $\{\mathsf{d},\mathsf{e}\}$.

8 Efficiency Considerations and the Specificity Ordering CP2

The specificity relations P1, P2, P3, and CP1[38] share several efficiency features, which we will highlight in this section. Moreover, we will introduce the specificity ordering CP2, a minor variation of CP1 toward more efficiency and intuitive adequacy. Finally, we will discuss further steps toward more efficiency following HERBRAND's Fundamental Theorem.

8.1 A Slight Gain in Efficiency

A straightforward procedure toward deciding the specificity relations $\lesssim_{\mathrm{CP1}}$ and $\lesssim_{\mathrm{P3}}$ between two arguments is to consider all possible activation sets from the literals in the sets $\mathfrak{T}_{\Pi}$ and $\mathfrak{T}_{\Pi\cup\Delta}$, respectively. The effort for computing $\lesssim_{\mathrm{CP1}}$ is lower than that of $\lesssim_{\mathrm{P3}}$ because of $\mathfrak{T}_{\Pi} \subseteq \mathfrak{T}_{\Pi\cup\Delta}$, though not w.r.t. asymptotic complexity: In both cases already the number of possible (simplified) activation sets is exponential in the number of literals in the respective sets $\mathfrak{T}_{\Pi}$ and $\mathfrak{T}_{\Pi\cup\Delta}$, because each possible subset has to be tested in principle.

8.2 Comparing Derivations

To lower the computational complexity, more syntactic criteria for computing specificity were introduced in [STOLZENBURG &AL., 2003]. These criteria refer to the *derivations* for the given arguments. More precisely, they refer to the *and-trees* of Definition 4.1 in § 4.4.1.

8.2.1 No Pruning Required

The concept of pruning and-trees is introduced in [STOLZENBURG &AL., 2003, Definition 12] in this context, because, for the case of $\lesssim_{\mathrm{P2}}$, attention cannot be restricted to derivations which make use only of the instances of defeasible rules given in the arguments. The reason for this is that the specificity notions according to [POOLE, 1985] and [SIMARI & LOUI, 1992] admit literals L in activation sets that cannot be derived solely by strict rules, i.e. $L \in \mathfrak{T}_{\Pi\cup\Delta} \setminus \mathfrak{T}_{\Pi}$. Since this is not possible with the relation $\lesssim_{\mathrm{CP1}}$, this problem vanishes with our corrected version of specificity. This problem and its vanishing are discussed in Example 7.4. of § 7.2.

8.2.2 Sets of Derivations have to be Compared

Yet still, the specificity relation $\lesssim_{\mathrm{CP1}}$ inherits several properties from $\lesssim_{\mathrm{P3}}$. For instance, the syntactic criteria of their definitions require us in general to compare two *sets* of derivations *element by element*. This is true for both specificity relations, as shown in the following example.

---

[37] Cf. Note 3 of § 2.3.

[38] P1 follows [POOLE, 1985] and can be found in this paper in Definition 6.3 of § 6.2. P2 follows [SIMARI & LOUI, 1992] and can be found in Definition 6.3 of § 6.2. P3 respects non-defeasible arguments and can be found in Definition 6.6 of § 6.2. CP1 is our transitive relation found in Definition 6.12 of § 6.4.

Two and-trees can be compared w.r.t. $\trianglelefteq$ efficiently. This requires the subset comparison of all paths of the two trees, respectively. Hence, the respective complexity is polynomial, at most $\mathrm{O}(n^3)$, where n is the overall number of nodes in the and-trees.* This made the relation $\trianglelefteq$ attractive for practical use in the context of [STOLZENBURG &AL., 2003] compared to the exponential comparison mention in § 8.1. As stated in the following definition, for a comparison of specificity we have to consider all and-trees, however, and so we still remain with an overall exponential time complexity, which is worse than the one we will describe in Remark 8.28 in § 8.3.4.

[why worse? both are exponential]

Definition 8.4 ([STOLZENBURG &AL., 2003, Definition 24])
$(\mathcal{A}_1, h_1) \leq (\mathcal{A}_2, h_2)$ if $(\mathcal{A}_1, h_1)$ and $(\mathcal{A}_2, h_2)$ are two arguments in the given specification and for each and-tree T_1 for h_1 there is an and-tree T_2 for h_2 such that $T_1 \trianglelefteq T_2$.

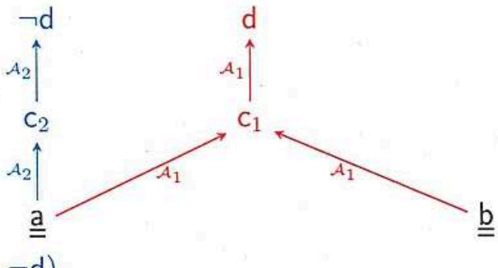
It is shown in [STOLZENBURG &AL., 2003, Theorem 25] that $\leq$ and $\precsim_{\mathrm{P2}}$ are equal in special cases, namely if the arguments involved in the comparison correspond to exactly one and-tree. Let us try to adapt this result to our new relation $\precsim_{\mathrm{CP1}}$, in the sense that we try to establish a mutual subset relation between $\leq$ and $\precsim_{\mathrm{CP1}}$.

The forward direction is pretty straightforward, but comes with the restriction to be expected: From [STOLZENBURG &AL., 2003, Theorem 25] we get $\leq\ \subseteq\ \precsim_{\mathrm{P2}}$. By looking at the empty path, we easily see that $\leq$ satisfies the additional restriction of Definition 6.6 as compared to Definition 6.4; so we also get $\leq\ \subseteq\ \precsim_{\mathrm{P3}}$. Finally, we can apply Theorem 6.16 and get the intended $\leq\ \subseteq\ \precsim_{\mathrm{CP1}}$, but only with the strong restriction of the condition of Theorem 6.16. We see no way yet to relax this restriction resulting from phase 3 of § 6.1.

It is even more unfortunate that the backward direction does not hold at all because of our change in phase 1 of § 6.1. In particular, as shown in the following example, it does not hold for the special case where it holds for $\precsim_{\mathrm{P2}}$, i.e. in the case that there are no general rules and hence each minimal argument corresponds to exactly one derivation (cf. the proof of Theorem 25 in [STOLZENBURG &AL., 2003]).

Example 8.5

$$\begin{aligned}
\Pi^{\mathrm{F}}_{8.5} &:= \{\mathsf{a}, \mathsf{b}\}, \quad \Pi^{\mathrm{G}}_{8.5} := \emptyset,\\
\Delta_{8.5} &:= \mathcal{A}_1 \cup \mathcal{A}_2.\\
\mathcal{A}_1 &:= \left\{\begin{array}{l} \mathsf{c}_1 \leftarrow \mathsf{a}\wedge\mathsf{b},\\ \mathsf{d} \leftarrow \mathsf{c}_1 \end{array}\right\}.\\
\mathcal{A}_2 &:= \left\{\begin{array}{l} \mathsf{c}_2 \leftarrow \mathsf{a},\\ \neg\mathsf{d} \leftarrow \mathsf{c}_2 \end{array}\right\}.
\end{aligned}$$



We have $(\mathcal{A}_1, \mathsf{d}) \vartriangle_{\mathrm{P3}} (\mathcal{A}_2, \neg\mathsf{d})$ and $(\mathcal{A}_1, \mathsf{d}) <_{\mathrm{CP1}} (\mathcal{A}_2, \neg\mathsf{d})$.

Both arguments $(\mathcal{A}_1, \mathsf{d})$ and $(\mathcal{A}_2, \neg\mathsf{d})$ correspond to exactly one and-tree, say T_1 and T_2, respectively. All paths in $\mathsf{Paths}(T_1)$ contain c_1, but not c_2, and all paths in $\mathsf{Paths}(T_2)$ contain c_2, but not c_1. Hence, $(\mathcal{A}_1, \mathsf{d}) \leq (\mathcal{A}_2, \neg\mathsf{d})$ does *not* hold.

* In order to achieve this complexity $O(n^3)$, the literals in the paths of the and-trees can be sorted. This makes comparing paths wrt. $\subseteq$ more easy.

As a step toward a more efficiently realizable notion of POOLE-style specificity, we will now eliminate those activation sets from our considerations that depend on and-trees with an inessential application of the instance of a defeasible rule.[40]

As a side effect, this step will also eliminate all redundant activation sets that result from what was called "growth of the defeasible parts toward the leaves" in § 4.4.3. This growth results from inessential application not of defeasible rules, but of general rules only. Contrary to the inessential application of instances of defeasible rules, this elimination of inessential applications of general rules will not change our specificity relation.

The positive effect, however, of cutting off this growth is the following: When the leaves of the defeasible part of an and-tree are included in $\mathfrak{T}_{\hat{\Pi}}$ for the first time in a root-to-leaves traversal, we *immediately* stop and obtain one single immediate activation set, and that's it! The further enumeration of subsumed activation sets is no longer required.

While this reduction of the number of activation sets to one single immediate activation set for each and-tree is most helpful for the computation related to the first argument of the relation $\lesssim_{\mathrm{CP2}}$ when trying to decide it, for the computation related to the second argument it re-introduces the complication we already had in our first sketch of a notion of specificity in § 4.3.2, as compared to the simplified, second version of this sketch in § 4.4.4, which was the basis for our first formal definition of activation sets in Definition 6.1 of § 6.1.

This complication is only a notational one. It requires the notion of *weakly* immediate activation sets in addition to (non-weakly) immediate ones. This complication does not mean any extra-computation, not even for the second argument in the test for $\lesssim_{\mathrm{CP2}}$: It is just so that the test whether every activation set of the first argument is subsumed by some activation set for the second argument becomes independent from the computation of activation sets. This independence has the advantage that we can optimize it in several directions: First of all, we must omit all rules from Π^{F} and Δ, which play some minor rôles in the computation of non-immediate activation sets (namely Π^{F} for acceptance as an activation set, and the instances of Δ that form the first element of the argument for expansion of activation sets). More important, however, ~~is~~ that we may also add some forward reasoning from the activation set computed for the first argument in the test for $\lesssim_{\mathrm{CP2}}$.

All in all, this means for our operationalization that the computation of activation sets (cf. Definition 6.1) has to be replaced with the computation of *immediate* activation sets according to the following definition, which also mirrors our isolation of defeasible parts of derivations in § 4.4.1 more directly than before, namely in the sense that a growth towards the leaves is avoided and the further dissection described in Note 7 of § 4.4.2 takes place.

It may be helpful for an intuitive understanding of the following definition to have a look at Figure 1 in § 6.1: The root tree depicted there is captured in item 2 of the following definition, its sub-trees in item 1.

[40] The first idea could be to take only activation sets all of whose literals occur in the condition of a rule in $\mathcal{A}$, for the respective argument $(\mathcal{A}, L)$. This idea, however, is too restrictive because also general rules may play a rôle in the defeasible parts of the derivations, cf. § 4.4.1.

Remark 8.9 (Difference between an Activation Set and an Immediate One)
An immediate activation set differs from an activation set in that certain defeasible parts may no longer participate in the derivation, namely those parts that derive a node labeled with an element of $\mathfrak{T}_{\hat{\Pi}}$.

This means that those deviations which contain inessential[43] applications of instances of defeasible rules can no longer increase the number of activation sets, i.e. can no longer reduce the specificity of an argument. Such a reduction may sound counter-intuitive, but ~~Note that such a reduction is against common intuition. Indeed,~~ we cannot see any reason why the fact that the first element of the argument may also be re-used to re-derive a literal of $\mathfrak{T}_{\hat{\Pi}}$ from $\mathfrak{T}_{\hat{\Pi}}$ should be relevant for the specificity of the argument.

Definition 8.10 (2nd Version of our Specificity Relation: $\precsim_{\mathrm{CP2}}$)
$(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP2}} (\mathcal{A}_2, L_2)$ iff $(\mathcal{A}_1, L_1)$ and $(\mathcal{A}_2, L_2)$ are arguments, and we have

1. $L_1 \in \mathfrak{T}_{\hat{\Pi}}$ or
2. $L_2 \notin \mathfrak{T}_{\hat{\Pi}}$ and every $H \subseteq \mathfrak{T}_{\hat{\Pi}}$ that is an [minimal] immediate activation set for $(\mathcal{A}_1, L_1)$ is a weakly immediate activation set for $(\mathcal{A}_2, L_2)$.

To see that nothing essential has changed, compare the following Corollary 8.11 to Corollary 6.13 of § 6.4.

Corollary 8.11 *If $(\mathcal{A}_1, L_1)$, $(\mathcal{A}_2, L_2)$ are arguments with $\mathcal{A}_1 \subseteq \mathcal{A}_2$, then any of the following conditions is sufficient for $(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP2}} (\mathcal{A}_2, L_2)$:*

1. *$L_1 = L_2$.*
2. *$L_2 \in \mathfrak{T}_{\hat{\Pi}} \Rightarrow L_1 \in \mathfrak{T}_{\hat{\Pi}}$ and $\{L_1\} \cup \hat{\Pi} \vdash \{L_2\}$.*
3. *$L_1 \in \mathfrak{T}_{\hat{\Pi}}$ (which is implied by $\mathcal{A}_1 = \emptyset$ by Definition 2.6).*

Remark 8.12 (Optional Minimality Restriction has No Effect)
Note that the omission of the optional restriction to *minimal* immediate activation sets for $(\mathcal{A}_1, L_1)$ in Definition 8.10 has no effect on the extension of the defined notion.

Proof: Suppose that $L_1, L_2 \notin \mathfrak{T}_{\hat{\Pi}}$, and that H'' is an immediate activation set for $(\mathcal{A}_1, L_1)$. Because the related derivation is finite, we may assume that H'' is finite w.l.o.g. Thus, there is a minimal immediate activation set $H \subseteq H''$ for $(\mathcal{A}_1, L_1)$. If we now assume $(\mathcal{A}_1, L_1) \precsim_{\mathrm{CP2}} (\mathcal{A}_2, L_2)$ with respect to a definition with the optional minimality restriction, then H is a weakly immediate activation set for $(\mathcal{A}_2, L_2)$, i.e. there is an immediate activation set $H' \subseteq \mathfrak{T}_{H \cup \hat{\Pi}^{\mathrm{G}}}$ for $(\mathcal{A}_2, L_2)$, which (because of the monotonicity of our logic) implies $H' \subseteq \mathfrak{T}_{H'' \cup \hat{\Pi}^{\mathrm{G}}}$, i.e. H'' is a weakly immediate activation set for $(\mathcal{A}_2, L_2)$ as well, as was to be shown.

---

[43] I.e. inessential in the sense of Definition 8.6. [maybe better inline]

Furthermore, we have

$$(\mathcal{A}_1, \mathsf{drink}) <_{\mathrm{CP1}} (\mathcal{A}_3, \mathsf{alarm}) \vartriangle_{\mathrm{CP2}} (\mathcal{A}_1, \mathsf{drink}):$$

The minimal [immediate] activation set $\{\mathsf{danger}\}$ for $(\mathcal{A}_3, \mathsf{alarm})$ is not an activation set for $(\mathcal{A}_1, \mathsf{drink})$. The only [immediate] activation set for $(\mathcal{A}_1, \mathsf{drink})$ that is a subset of $\mathfrak{T}_{\hat{\Pi}_{8.15}}$ is $\{\mathsf{thirst}\}$, which is an activation set for $(\mathcal{A}_3, \mathsf{alarm})$, *but not a weakly immediate one.* Note that $(\mathcal{A}_1, \mathsf{drink})$ is no longer more or equivalently specific than $(\mathcal{A}_3, \mathsf{alarm})$ in the sense of $(\mathcal{A}_1, \mathsf{drink}) \not\lesssim_{\mathrm{CP2}} (\mathcal{A}_3, \mathsf{alarm})$, because the inessential application of the rule $\mathsf{danger}{\leftarrow}\mathsf{thirst}$ of $\mathcal{A}_3$ is not admitted for *immediate* activation sets.

In spite of these minor but noticeable differences, however, nothing has actually changed by stepping from CP1 to CP2, except the positioning of the argument $(\mathcal{A}_3, \mathsf{alarm})$, which is non-minimal as an argument (and therefore practically irrelevant and not even considered in many frameworks, cf. Remark 2.8 of § 2.4) and also non-minimal in $\lesssim_{\mathrm{CP1}}$ (and therefore less specific and not really relevant either). What is crucial, however, is that a most specific argument cannot be found in either case. Indeed, we have both

$$(\mathcal{A}_1, \mathsf{drink}) \vartriangle_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{alarm})$$

and $$(\mathcal{A}_1, \mathsf{drink}) \vartriangle_{\mathrm{CP2}} (\mathcal{A}_2, \mathsf{alarm}).$$

If we remove danger from $\Pi^{\mathrm{F}}_{8.15}$, then $(\mathcal{A}_2, \mathsf{alarm})$ is no argument anymore, but we can embed the specification injectively into the one of Example 3.3 of § 3 and get both

$$(\mathcal{A}_1, \mathsf{drink}) \approx_{\mathrm{CP1}} (\mathcal{A}_3, \mathsf{alarm})$$

and $$(\mathcal{A}_1, \mathsf{drink}) \approx_{\mathrm{CP2}} (\mathcal{A}_3, \mathsf{alarm}),$$

because the activation set $\{\mathsf{thirst}\}$ now becomes an immediate one also for $(\mathcal{A}_3, \mathsf{alarm})$. Indeed, the application of the rule $\mathsf{danger}{\leftarrow}\mathsf{thirst}$ is no longer inessential for deriving alarm.

Moreover, if we now add the rule $\mathsf{danger}{\Leftarrow}\mathsf{thirst}$ to $\Pi^{\mathrm{G}}_{8.15}$, resulting in the specification $(\{\mathsf{thirst}\}, \{\mathsf{danger}{\Leftarrow}\mathsf{thirst}\}, \Delta_{8.15})$, then the situation is essentially the same as in Example 3.2 of § 3, and so we get both $(\mathcal{A}_1, \mathsf{drink}) <_{\mathrm{CP1}} (\mathcal{A}_3, \mathsf{alarm}) \approx_{\mathrm{CP1}} (\mathcal{A}_2, \mathsf{alarm})$

and $(\mathcal{A}_1, \mathsf{drink}) <_{\mathrm{CP2}} (\mathcal{A}_3, \mathsf{alarm}) \approx_{\mathrm{CP2}} (\mathcal{A}_2, \mathsf{alarm})$,

because — although the application of the rule $\mathsf{danger}{\leftarrow}\mathsf{thirst}$ becomes inessential again by $\mathsf{danger} \in \mathfrak{T}_{\hat{\Pi}}$ — $\{\mathsf{thirst}\}$ now becomes a weakly immediate activation set for $(\mathcal{A}_3, \mathsf{alarm})$ and for $(\mathcal{A}_2, \mathsf{alarm})$, though not a (non-weakly) immediate one.

Corollary 8.16 ($\lesssim_{\mathrm{CP1}}$ and $\lesssim_{\mathrm{CP2}}$ are incomparable)
There are a specification $(\Pi^{\mathrm{F}}_{8.15}, \Pi^{\mathrm{G}}_{8.15}, \Delta_{8.15})$ (without any negative literals) and arguments $(\mathcal{A}_1, L_1)$, $(\mathcal{A}_3, L_3)$, $(\mathcal{A}_2, L_2)$, such that $(\mathcal{A}_1, L_1) \lesssim_{\mathrm{CP1}} (\mathcal{A}_3, L_3) \lesssim_{\mathrm{CP2}} (\mathcal{A}_2, L_2)$

and $(\mathcal{A}_1, L_1) \not\lesssim_{\mathrm{CP2}} (\mathcal{A}_3, L_3) \not\lesssim_{\mathrm{CP1}} (\mathcal{A}_2, L_2)$,

i.e. $\lesssim_{\mathrm{CP1}} \not\subseteq \lesssim_{\mathrm{CP2}} \not\subseteq \lesssim_{\mathrm{CP1}}$.

Nevertheless, Example 8.15 suggests that only some unimportant details make $\lesssim_{\mathrm{CP1}}$ and $\lesssim_{\mathrm{CP2}}$ incomparable to each other, but that the most specific minimal arguments seem to remain most specific and so nothing essential seems to change.

So we may ask ourselves: What changes occur in our previous examples when we switch from CP1 to CP2? Do any of the relations stated for CP1 change for CP2?

The answer to the latter question is: No!

We would like to ask the reader to check this carefully.

Proof of Lemma 8.19
Let $(\mathcal{A}, L)$ be a minimal argument.

In case of $L \in \mathfrak{T}_{\hat{\Pi}}$, there is exactly one minimal activation set for $(\mathcal{A}, L)$, namely the empty set $\emptyset$, and this an immediate activation set (choose $\mathfrak{L} := \emptyset$ in Definition 8.7). Moreover, because $(\mathcal{A}, L)$ is a minimal argument, we have $\mathcal{A} = \emptyset$, and then $\mathfrak{C} = \emptyset$. So we get our unique minimal activation set $\emptyset$ indeed in the claimed form of $\mathfrak{C} \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}} = \emptyset \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}} = \emptyset$.

can may be done simpler, because minim. arguments, only one and-tree

It now remains to consider the case of $L \notin \mathfrak{T}_{\hat{\Pi}}$. Because $(\mathcal{A}, L)$ is an argument, there is an and-tree for the derivation of $\hat{\Pi}^{\mathrm{F}} \cup \mathcal{A} \cup \hat{\Pi}^{\mathrm{G}} \vdash \{L\}$. As every and-tree is finite, there is a finite activation set $H' \subseteq \hat{\Pi}^{\mathrm{F}}$ for $(\mathcal{A}, L)$. Then there must be a minimal activation set H for $(\mathcal{A}, L)$ with $H \subseteq H'$. Then we have $H \subseteq \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}}$. Then there is an and-tree T for the derivation of $H \cup \mathcal{A} \cup \hat{\Pi}^{\mathrm{G}} \vdash \{L\}$ (which is actually unique, but this does not matter here). Let $\mathfrak{D}$ be the set of all conclusions of all rules in $\mathcal{A}$. Let $\mathfrak{D}'$ be the set of all literals in $\mathcal{A}$ (i.e. rules with empty conditions). Then $\mathfrak{D}' \subseteq \mathfrak{D}$. Because $(\mathcal{A}, L)$ is a minimal argument, we know that $\mathfrak{D} \cap \mathfrak{T}_{\hat{\Pi}} = \emptyset$ and that every rule from $\mathcal{A}$ is applied in T. Thus, because of $L \notin \mathfrak{T}_{\hat{\Pi}}$ and because all rules in $\hat{\Pi}$ are just literals, the set of the labels of the leaves of T is exactly $(\mathfrak{C} \cap \mathfrak{T}_{\hat{\Pi}}) \cup \mathfrak{D}'$. Because T is an and-tree for the derivation of $H \cup \mathcal{A} \cup \hat{\Pi}^{\mathrm{G}} \vdash \{L\}$, because $\mathcal{A} \cap \mathfrak{T}_{\hat{\Pi}} \subseteq \mathfrak{D}' \cap \mathfrak{T}_{\hat{\Pi}} \subseteq \mathfrak{D} \cap \mathfrak{T}_{\hat{\Pi}} = \emptyset$, and because all rules in $\hat{\Pi}^{\mathrm{G}}$ are just literals, we have $\mathfrak{C} \cap \mathfrak{T}_{\hat{\Pi}} \subseteq (H \cup \mathcal{A} \cup \hat{\Pi}^{\mathrm{G}}) \cap \mathfrak{T}_{\hat{\Pi}} = H \cup \emptyset \cup \hat{\Pi}^{\mathrm{G}} = H \cup \hat{\Pi}^{\mathrm{G}}$, $\mathfrak{T}_{\hat{\Pi}^{\mathrm{G}}} = \hat{\Pi}^{\mathrm{G}}$, and $\mathfrak{T}_{\hat{\Pi}} = \hat{\Pi}^{\mathrm{F}} \cup \hat{\Pi}^{\mathrm{G}}$. Because H is a *minimal* activation set for $(\mathcal{A}, L)$, H must be a subset of the leaves of T not in $\mathfrak{D}'$: $H \subseteq \mathfrak{C} \cap \mathfrak{T}_{\hat{\Pi}}$. Because of our previous result of $H \subseteq \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}}$, we now get by our two subset properties $H \subseteq \mathfrak{C} \cap \mathfrak{T}_{\hat{\Pi}} \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}} \subseteq (H \cup \hat{\Pi}^{\mathrm{G}}) \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}} = H \cup \emptyset = H$, i.e. $H = \mathfrak{C} \cap \mathfrak{T}_{\hat{\Pi}} \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}} = \mathfrak{C} \cap (\hat{\Pi}^{\mathrm{F}} \cup \hat{\Pi}^{\mathrm{G}}) \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}} = \mathfrak{C} \cap \hat{\Pi}^{\mathrm{F}} \backslash \hat{\Pi}^{\mathrm{G}}$. Choosing $\mathfrak{L} := \{L\}$ in item 1 and a proof tree consisting only of a root in item 2 of Definition 8.7, we see that H is actually an *immediate* activation set for $(\mathcal{A}, L)$; in particular we have $L \notin \mathfrak{T}_{\hat{\Pi}}$ and the property required in the last line of item 1 of Definition 8.7: $(\mathfrak{C} \cap \mathfrak{T}_{\hat{\Pi}}) \cup \mathfrak{D}' \subseteq H \cup \mathfrak{T}_{\hat{\Pi}} \cup \mathcal{A}$. Finally, H is a *minimal* immediate activation set by Corollary 8.8(5). **Q.e.d. (Lemma 8.19)**

The second trivial form of classification is to take all rules without conditions to be defeasible. It is not a good idea for comparing arguments w.r.t. specificity, however:

Corollary 8.20 *Assume that $\Pi^{\mathrm{F}} = \emptyset$ and that Π^{G} contains only rules with non-empty conditions. Then we have $\mathfrak{T}_{\hat{\Pi}} = \emptyset$. Moreover, for every argument, there is exactly one [immediate] activation set H with $H \subseteq \mathfrak{T}_{\hat{\Pi}}$, namely $H = \emptyset$. Furthermore, all arguments are equivalent w.r.t. $\approx_{\mathrm{CP1}}$ and $\approx_{\mathrm{CP2}}$.*

Finally, note that the computation of simplified activation sets that are a subset of $\mathfrak{T}_{\hat{\Pi} \cup \hat{\Delta}}$ — as required for P1, P2, P3 instead of CP1, CP2 — is not simplified for the special cases of this section, contrary to the computation of [immediate] activation sets that are subsets of $\mathfrak{T}_{\hat{\Pi}}$.

procedure immediate-activation-sets(L):
(* L must be a literal *)
if $L \notin \mathfrak{T}_{\hat{\Pi}}$ then (call immediate-activation-sets-helper($\{(L, 2)\}, \emptyset, \emptyset, \emptyset, L$)).

procedure immediate-activation-sets-helper(T, O, H, A, I):
(* T is the current goal. T must be a set of pairs (L, B) of a literal $L \notin \mathfrak{T}_{\hat{\Pi}}$ and
a bit $B \in \{1, 2\}$ referring to the two items of Definition 8.7,
such that $B = 1$ indicates that L labels a defeasible part *)
(* O is a set of literals that indicate that our algorithm may have missed
to enumerate a most general immediate activation set in case of $O \cap \mathfrak{T}_{\hat{\Pi}} \neq \emptyset$
because the and-tree has already been properly expanded at their nodes
(which occur in a defeasible part!) *)
(* H is an accumulator for the immediate activation set,
H must always be a set of literals $L \in \mathfrak{T}_{\hat{\Pi}}$ from the fringes of defeasible parts *)
(* A is an accumulator for the first element of the argument *)
(* I is the possibly instantiated input literal and second element of the argument *)
if $T = \emptyset$ then (output "`H is immediate activation set for` (A, I)" and exit);
pick some (L, B) from T; $T := T \setminus \{(L, B)\}$;
for each rule $(L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n) \in \Pi \cup \Delta$ do
for some ξ that maps all variables in $L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n$ to fresh variables do
if L and $L'\xi$ have the most general unifier σ then [
$I' := I\sigma$; if $I' \in \mathfrak{T}_{\hat{\Pi}}$ then (output "`Instance` $I' \in \mathfrak{T}_{\hat{\Pi}}$" and exit);
$O' := O\sigma$; if $O' \cap \mathfrak{T}_{\hat{\Pi}} \neq \emptyset$ then (output "`breach`" and exit);
$T' := \{ (L'''\sigma, B''') \mid (L''', B''') \in T \wedge L'''\sigma \notin \mathfrak{T}_{\hat{\Pi}} \}$;
$H' := H\sigma \cup \{ L'''\sigma \mid (L''', 1) \in T \wedge L'''\sigma \in \mathfrak{T}_{\hat{\Pi}} \}$;
$A' := A\sigma$;
if $L\sigma \in \mathfrak{T}_{\hat{\Pi}}$ then (if $B = 1$ then $(H' := H' \cup \{L\sigma\})$)
else (
$B' := B$;
if $(L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n) \notin \Pi$ then (
(* The applied rule is necessarily a defeasible one! *)
$A' := A' \cup \{(L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n)\xi\sigma\}$;
$B' := 1$);
$T' := T' \cup \{ (L''_i\xi\sigma, B') \mid i \in \{1, \ldots, n\} \wedge L''_i\xi\sigma \notin \mathfrak{T}_{\hat{\Pi}} \}$;
if $B' = 1 \wedge n \geq 1$ then (
(* $B' = 1$ means that we are in a defeasible part now,
and so we have to accumulate our activation set! *)
(* $n \geq 1$ means that we have to expand the and-tree *properly*
under the crucial assumption that $L\sigma \notin \mathfrak{T}_{\hat{\Pi}}$. *)
$H' := H' \cup \{ L''_i\xi\sigma \mid i \in \{1, \ldots, n\} \wedge L''_i\xi\sigma \in \mathfrak{T}_{\hat{\Pi}} \}$;
$O' := O' \cup \{L\sigma\}$));
$O' := \{ L''' \in O' \mid L'''$ is not ground $\}$;
call immediate-activation-sets-helper(T', O', H', A', I')].

Figure 2: Sketch of immediate-activation-sets and immediate-activation-sets-helper

[Mention that Prolog implementation ohay exists and is available by the authors] p.55

one, namely $\mathsf{flies}(x)\leftarrow\mathsf{bird}(x)$ from $\Delta_{3.3}$. We can take ξ and σ as the identity and $\{x\mapsto y\}$, respectively. The program variable B' will be set to 1, and the tail-recursive call will have the argument tuple

$$(\ \{(\mathsf{bird}(y),1)\},\ \ \{\mathsf{flies}(y)\},\ \ \emptyset,\ \ \{\mathsf{flies}(y)\leftarrow\mathsf{bird}(y)\},\ \ \mathsf{flies}(y)\ ).$$

Again the only rule whose conclusion is unifiable with the only goal literal is a defeasible one, namely $\mathsf{bird}(x)\leftarrow\mathsf{emu}(x)$ from $\Delta_{3.3}$. We can again take ξ and σ as the identity and $\{x\mapsto y\}$, respectively. The program variable B' will be set to 1, and the tail-recursive call will have the argument tuple

$$(\ \{(\mathsf{emu}(y),1)\},\ \ \{\mathsf{flies}(y),\mathsf{bird}(y)\},\ \ \emptyset,\ \ \{\mathsf{flies}(y)\leftarrow\mathsf{bird}(y),\mathsf{bird}(y)\leftarrow\mathsf{emu}(y)\},\ \ \mathsf{flies}(y)\ ).$$

Now the only rule whose conclusion is unifiable with the only goal literal is a fact, namely $\mathsf{emu}(\mathsf{edna})$ from $\Pi^{\mathrm{F}}_{3.3}$. We can take ξ and σ as the identity and $\{y\mapsto\mathsf{edna}\}$, respectively. The program variable B' will be set to 1, and the tail-recursive call will have the argument tuple

$$(\ \emptyset,\ \ \emptyset,\ \ \{\mathsf{emu}(\mathsf{edna})\},\ \ \{\mathsf{flies}(\mathsf{edna})\leftarrow\mathsf{bird}(\mathsf{edna}),\mathsf{bird}(\mathsf{edna})\leftarrow\mathsf{emu}(\mathsf{edna})\},\ \ \mathsf{flies}(\mathsf{edna})\ ).$$

This call immediately terminates by outputting the immediate activation set $\{\mathsf{emu}(\mathsf{edna})\}$ for the argument $(\ \{\mathsf{flies}(\mathsf{edna})\leftarrow\mathsf{bird}(\mathsf{edna}),\mathsf{bird}(\mathsf{edna})\leftarrow\mathsf{emu}(\mathsf{edna})\},\ \ \mathsf{flies}(\mathsf{edna})\ )$. As all calls are terminated now and there was no output `"breach"`, this means that we have enumerated all immediate activation sets for all instances of the input literal.

Example 8.25 *(continuing Example 3.2 of § 3)*

Let us now come to Example 3.2 of § 3. We start with the same input as for Example 8.24 above, and there is no change up to the call with argument tuple

$$(\ \{(\mathsf{bird}(y),1)\},\ \ \{\mathsf{flies}(y)\},\ \ \emptyset,\ \ \{\mathsf{flies}(y)\leftarrow\mathsf{bird}(y)\},\ \ \mathsf{flies}(y)\ ),$$

and the only difference before the next call is that the applied rule is a strict one and is not recorded in the program variable A'. Thus, we get a call with the argument tuple

$$(\ \{(\mathsf{emu}(y),1)\},\ \ \{\mathsf{flies}(y),\mathsf{bird}(y)\},\ \ \emptyset,\ \ \{\mathsf{flies}(y)\leftarrow\mathsf{bird}(y)\},\ \ \mathsf{flies}(y)\ ).$$

There is still no essential change, except that the test for `"breach"` becomes positive: We again have $O\sigma=\{\mathsf{flies}(\mathsf{edna}),\mathsf{bird}(\mathsf{edna})\}$, but now we have $\mathsf{bird}(\mathsf{edna})\in\mathfrak{T}_{\hat\Pi}$, and our procedure outputs `"breach"`. Indeed, it missed to enumerate the immediate activation set $\{\mathsf{bird}(\mathsf{edna})\}$ for the argument $(\{\mathsf{flies}(\mathsf{edna})\leftarrow\mathsf{bird}(\mathsf{edna})\},\mathsf{flies}(\mathsf{edna}))$, simply because the instantiation came too late to stop us from proper expansion of the and-tree.

Remark 8.26 (Closer Matching of Activation Sets to SLD-Resolution Results in Inappropriate Semantics)

The obvious idea to avoid the possibility that the procedure of Figure 2 may output `"breach"` and miss some maximally general, immediate activation sets is the following.

Just like we obtained CP2 from CP1, it is possible to obtain a notion CP3 from CP2 by a minor modification of immediate activation sets, resulting in, say, *SLD activation sets*, such that the SLD-like computation of Figure 2 enumerates all maximally general, SLD activation sets.

We do not see a chance to satisfy the crucial requirement of such a modification, however, namely that it does not affect any of our previous examples. If we look at the application of the procedure of Figure 2 to the specification of Example 3.2 as described in Example 8.25, then we see that all SLD activation sets remaining in Example 3.2 could be $\{\mathsf{emu}(\mathsf{edna})\}$, such that the arguments $(\mathcal{A}_1,\neg\mathsf{flies}(\mathsf{edna}))$ and $(\mathcal{A}_2,\mathsf{flies}(\mathsf{edna}))$ would become equivalently specific w.r.t. the specification of Example 3.2, which seems to be absurd.

Remark 8.27 (Specificity Relation on Arguments Extended with an And-Tree) A straightforward idea to improve tractability is to attach an and-tree to each argument and to compute a unique[52] immediate activation set for each argument as follows:

Starting from the root, we traverse the tree, remembering whether we have passed an application of the instance of a defeasible rule, and stop traversing at the first node labeled with an element of the finite set $\mathfrak{T}_{\Pi_{\mathrm{G}}}$, outputting its literal as part of the single *tree-immediate activation set*, provided that we have passed an application of the instance of a defeasible rule.

The problem we see here, however, is that such a fixed and-tree does not make much sense for the second argument of our relation $\lesssim_{\mathrm{CP2}}$, simply because we should not let an inappropriately chosen and-tree for the second argument produce a failure of the property of being more specific. This means that we need an existential quantification over the and-tree of the second argument. If we were able to find a way to handle this quantification, the same technique would probably admit us to handle a universal quantification over the and-tree of the first argument, which brings us back to our original relation $\lesssim_{\mathrm{CP2}}$ on arguments without and-trees.

So this restriction to concrete and-trees does not seem to help. We will now show that we do not need it either.

procedure ground-immediate-activation-sets-helper(T, H, A):
(* T is the current goal. T must be a set of pairs (L, B) of a literal $L \notin \mathfrak{T}_{\Pi_{\mathrm{G}}}$ and
  a bit $B \in \{1, 2\}$ referring to the two items of Definition 8.7,
  such that $B = 1$ indicates that L labels a defeasible part *)
(* H is an accumulator for the immediate activation set,
  H must always be a set of literals $L \in \mathfrak{T}_{\Pi_{\mathrm{G}}}$ from the fringes of defeasible parts *)
(* A is an accumulator for the first element of the argument *)
(* note that the input literal I is invariant now; no input, no output *)
  if $T = \emptyset$ then (output (H, A) and exit);
  pick some (L, B) from T; $T := T \setminus \{(L, B)\}$;
  (* We do not have to test rules from $\Pi^{\mathrm{F}}_{\mathrm{G}}$ because of $L \notin \mathfrak{T}_{\Pi_{\mathrm{G}}}$. *)
  for each rule $(L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n) \in \Pi^{\mathrm{G}}_{\mathrm{G}} \cup \Delta_{\mathrm{G}}$ do
  if $L = L'$ then [
    $H' := H$; $A' := A$; $B' := B$;
    if $(L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n) \notin \Pi^{\mathrm{G}}_{\mathrm{G}}$ then (
      (* The applied rule is now necessarily a defeasible one. *)
      $A' := A' \cup \{(L' \Leftarrow L''_1 \wedge \ldots \wedge L''_n)\}$;
      $B' := 1$);
    $T' := T \cup \{\ (L''_i, B') \mid i \in \{1, \ldots, n\} \wedge L''_i \notin \mathfrak{T}_{\Pi_{\mathrm{G}}}\ \}$;
    if $B' = 1$ then (
      (* $B' = 1$ means that we are in a defeasible part now,
        and so we have to accumulate our activation set! *)
      $H' := H' \cup \{\ L''_i \mid i \in \{1, \ldots, n\} \wedge L''_i \in \mathfrak{T}_{\Pi_{\mathrm{G}}}\ \}$);
    call ground-immediate-activation-sets-helper(T', H', A')].

Figure 3: Sketch of procedure ground-immediate-activation-sets-helper

[52] See, however, Example 8.18 in § 8.3.1.

9 Conclusion

the lack of transitivity

We would need further discussions on our surprising new findings — after all, defeasible reasoning with POOLE's notion of specificity is being applied now for over a quarter of century, and it was not to be expected that our investigations could shake an element of the field to the very foundations.

One remedy for the discovered lack of transitivity of $\lesssim_{\mathrm{P3}}$ could be to consider the transitive closure of the non-transitive relation $\lesssim_{\mathrm{P3}}$. This could be an advantage only under the condition that the transitive closure of $\lesssim_{\mathrm{P3}}$ is a subset of $\lesssim_{\mathrm{CP1}}$, i.e. only under the condition of Theorem 6.16. Moreover, this transitive closure will still have all the intuitive shortcomings made obvious in § 7. Furthermore, we do not see how this transitive closure could be decided efficiently. Finally, the transitive closure lacks a direct intuitive motivation, and after the first extension step from $\lesssim_{\mathrm{P3}}$ to its transitive closure, we had better take the second extension step to the more intuitive $\lesssim_{\mathrm{CP1}}$ immediately.

Contrary to the transitive closure of $\lesssim_{\mathrm{P3}}$, our novel relations $\lesssim_{\mathrm{CP1}}$ and $\lesssim_{\mathrm{CP2}}$ also solve the problem of non-monotonicity of specificity w.r.t. conjunction (cf. § 7.1), which was already realized as a problem of $\lesssim_{\mathrm{P1}}$ by [POOLE, 1985] (cf. Example 7.1).

The present means to decide our novel specificity relation $\lesssim_{\mathrm{CP1}}$, however, show several improvements[54] and a few setbacks[55] compared to the known ones for POOLE's relation. Further work is needed to improve efficiency.

By a minor restriction of activations sets, resulting in *immediate* activations sets, we have come in § 8.3 to the quasi-ordering $\lesssim_{\mathrm{CP2}}$, which does not show any difference compared to $\lesssim_{\mathrm{CP1}}$ in any of our examples except Example 8.15, which was constructed to show the difference. The new specificity ordering $\lesssim_{\mathrm{CP2}}$ has advantages w.r.t. intuition and efficiency. The latter advantage, however, requires decidability of $\mathfrak{T}_{\hat{\Pi}}$ (in addition to the always given semi-decidability). To concretize the problems of computing activation sets by SLD-resolution we have sketched a procedure that may indicate `"breach"` if it may have missed some output in § 8.3.3. Then, in § 8.3.4, we have shown how to obtain decidability of $\mathfrak{T}_{\hat{\Pi}}$ by restriction to a finite set of instances that are then treated as if they were ground. Without such a restriction we do not know how to decide any of the relations $\lesssim_{\mathrm{P1}}$, $\lesssim_{\mathrm{P2}}$, $\lesssim_{\mathrm{P3}}$, $\lesssim_{\mathrm{CP1}}$, $\lesssim_{\mathrm{CP2}}$; and we hope that we can find a procedure for generating the finite set of rule instances such that the effect of this restriction can be neglected in practice.

If we look beyond the very subject of this paper, we see that an important part of the application context consists of numerous frameworks for argumentation in logic. The overall process usually includes a dialectical process used for answering queries. Different arguments are pro or contra a certain answer. By means of an attack relation conflicts between contradicting arguments can be determined in abstract argumentation frameworks, such as [DUNG, 1995], [PRAKKEN & VREESWIJK, 2002], [MODGIL & PRAKKEN, 2014]. A concrete specificity or similar relation helps then to decide among conflicting arguments. As the discussion in this paper demonstrates, it is not that easy, however, to find an effective concrete specificity relation. One of the main problems is that such relations are often computationally highly complex (such as in [KERN-ISBERNER & THIMM, 2012]).

[54] See §§ 8.1, 8.2.1, 8.3.2, 8.3.3, and 8.3.4 for the improvements.

[55] See §§ 8.2.3 and 8.3.3 for the setbacks.

we presented good intuitive reasons for the failure of the preference of Example 3.3 in Example 6.20 of § 6.6 (cf. also the pointers to further reasons in Note 32 to Example 6.20).

It is just too early for a further conclusion, and the further implications of the contributions of this paper and the technical details of the operationalization of our correction of POOLE's specificity will have to be discussed in future work.

[Ratiolog project gives motivation for further work: natural-language question answering]

Acknowledgments

This research has been supported by the DFG (German Research Assoc.) grant Sto 421/5-1.

We would like to thank MATTHIAS THIMM for his helpful discussions of Example 3.3. Moreover, we are deeply obliged to several anonymous reviewers of previous versions of this paper for their most profound critiques and suggestions for improvement, which have been an enormous help to improve the quality of presentation considerably.

To honor DAVID POOLE, let us end this paper with the last sentence of [POOLE, 1985]:

This research was sponsored by no defence department.

References

[AIELLO &AL., 1996] Luigia Carlucci Aiello, Jon Doyle, and Stuart C. Shapiro, editors. *Proc. 5th KR 1996 — Int. Conf. on Principles of Knowledge Representation and Reasoning, Nov. 5–8, 1996, Cambridge (MA).* Morgan Kaufmann (Elsevier), Los Altos (CA), 1996.

[BARAL &AL., 2014] Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors. *Proc. 14th KR 2014 — Int. Conf. on Principles of Knowledge Representation and Reasoning, Jul 20–24, as part of the Vienna Summer of Logic, Vienna, July 9–24, 2014.* AAAI Press, 2014. ~~http://www.kr.org/index.php?page=proclist~~ [will probably not remain]

[BENFERHAT & GARCIA, 1997] Salem Benferhat and Laurent Garcia. A coherence-based approach to default reasoning. 1997. In [GABBAY &AL., 1997, pp. 43–57]. http://dx.doi.org/10.1007/BFb0035611. [crossreferences only if needed more than once, can be done with BibTeX]

[BESNARD & HUNTER, 2001] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001. Received Dec. 8, 2000. http://dx.doi.org/10.1016/S0004-3702(01)00071-6.

[56] Let us compare our specificity relations P3, CP1, CP2 with the "more conservative"-quasi-ordering by looking at the argument comparisons that were just made to clarify our notions, i.e. at our Corollaries 6.7, 6.13, and 8.11 in the context of Corollary 6.8. So let us assume $\mathcal{A}_1 \subseteq \mathcal{A}_2$. For the trivial case of $L_1 = L_2$, the argument $(\mathcal{A}_1, L_1)$ is quasi-smaller than the argument $(\mathcal{A}_2, L_2)$ for all of P3, CP1, CP2, and "more conservative". In case of $L_2 \in \mathfrak{T}_{\hat{\Pi}} \Rightarrow L_1 \in \mathfrak{T}_{\hat{\Pi}}$ and $\{L_1\} \cup \hat{\Pi} \vdash \{L_2\}$, again the argument $(\mathcal{A}_1, L_1)$ is quasi-smaller than the argument $(\mathcal{A}_2, L_2)$ for all of P3, CP1, CP2, but for "more conservative" it is the other way round, provided that we adopt the straightforward assumption that derivability is derivability w.r.t. the basic theory of $\hat{\Pi}$. Thus, P3, CP1, CP2 would all prefer $(\mathcal{A}, \mathsf{Mother}(\text{CLAIRE DE LONG}, \text{PIERRE FERMAT}))$ to $(\mathcal{A}, \exists x.\ \mathsf{Mother}(x, \text{PIERRE FERMAT}))$, provided that we could express existential quantification.

[HERBRAND, 1930] Jacques Herbrand. *Recherches sur la théorie de la démonstration.* PhD thesis, Université de Paris, 1930. Thèses présentées à la faculté des Sciences de Paris pour obtenir le grade de docteurès sciences mathématiques — 1re thèse: Recherches sur la théorie de la démonstration — 2me thèse: Propositions données par la faculté, Les équations de Fredholm — Soutenues le 1930 devant la commission d'examen — Président: M. VESSIOT, Examinateurs: MM. DENJOY, FRECHET — Vu et approuvé, Paris, le 20 Juin 1929, Le doyen de la faculté des Sciences, C. MAURAIN — Vu et permis d'imprimer, Paris, le 20 Juin 1929, Le recteur de l'Academie de Paris, S. CHARLETY — No. d'ordre 2121, Série A, No. de Série 1252 — Imprimerie J. Dziewulski, Varsovie — Univ. de Paris. [Also in Prace Towarzystwa Naukowego Warszawskiego, Wydział III Nauk Matematyczno-Fizychnych, Nr. 33, Warszawa. A contorted, newly typeset reprint is [HERBRAND, 1968, pp. 35–153]. Annotated English translation *Investigations in Proof Theory* by WARREN GOLDFARB (Chapters 1–4) and BURTON DREBEN and JEAN VAN HEIJENOORT (Chapter 5) with a brief introduction by GOLDFARB and extended notes by GOLDFARB (Notes A–C, K–M, O), DREBEN (Notes F–I), DREBEN and GOLDFARB (Notes D, J, and N), and DREBEN, GEORGE HUFF, and THEODORE HAILPERIN (Note E) in [HERBRAND, 1971, pp. 44–202]. English translation of § 5 with a different introduction by HEIJENOORT and some additional extended notes by DREBEN also in [HEIJENOORT, 1971, pp. 525–581]. *(HERBRAND's PhD thesis, his cardinal work, dated April 14, 1929; submitted at the Univ. of Paris; defended at the Sorbonne June 11, 1930; printed in Warsaw, 1930.)* [delete!]

[HERBRAND, 1968] Jacques Herbrand. *Écrits Logiques.* Presses Universitaires de France, Paris, 1968. Cortorted edn. of HERBRAND's logical writings by JEAN VAN HEIJENOORT. Review in [GOLDFARB, 1970]. English translation is [HERBRAND, 1971].

[HERBRAND, 1971] Jacques Herbrand. *Logical Writings.* Harvard Univ. Press, 1971. Ed. by WARREN GOLDFARB. Translation of [HERBRAND, 1968] with additional annotations, brief introductions, and extended notes by GOLDFARB, BURTON DREBEN, and JEAN VAN HEIJENOORT. [~~*(This edition is still an excellent source on* HERBRAND*'s writings today,*~~ *but it is problematic because it is based on the contorted* ~~*reprint*~~ *[*HERBRAND*, 1968]. This means that it urgently* ~~*needs a corrected edition*~~ *based on the original editions of* HERBRAND*'s* ~~*logical writings*~~*, which are all in French and which* ~~*should be included in facsimile to avoid future confusion.)*~~

[JOSHI, 1985] Aravind Joshi, editor. *Proc. 9th Int. Joint Conf. on Artificial Intelligence (IJCAI), 1985, Aug. 18–25, Los Angeles (CA).* Morgan Kaufmann (Elsevier), Los Altos (CA), 1985. `http://ijcai.org/Past%20Proceedings`.

[KERN-ISBERNER & THIMM, 2012] Gabriele Kern-Isberner and Matthias Thimm. A ranking semantics for first-order conditionals. 2012. [RAEDT &AL., 2012, pp. 456–461]. `http://dx.doi.org/10.3233/978-1-61499-098-7-456`.

[KOWALSKI, 1974] Robert A. Kowalski. Predicate logic as a programming language. 1974. In [ROSENFELD, 1974, pp. 569–574].

[LAMBERT, 1764] Johann Heinrich Lambert. *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung von*

[not needed]

~~Schmidt Str. 5, D-28359 Bremen,~~ Germany, 2013. ii+34 pp., http://arxiv.org/abs/1308.4943.

[WIRTH & STOLZENBURG, 2014] Claus-Peter Wirth and Frieder Stolzenburg. DAVID POOLE's specificity revised. 2014. In [BARAL &AL., 2014, pp. 168–177]. Short version of [WIRTH & STOLZENBURG, 2013].

[WIRTH &AL., 2009] Claus-Peter Wirth, Jörg Siekmann, Christoph Benzmüller, and Serge Autexier. JACQUES HERBRAND: Life, logic, and automated deduction. 2009. In [GABBAY & WOODS, 2004ff., Vol. 5: Logic from RUSSELL to CHURCH, pp. 195–254].

[WIRTH &AL., 2014] Claus-Peter Wirth, Jörg Siekmann, Christoph Benzmüller, and Serge Autexier. *Lectures on* JACQUES HERBRAND *as a Logician.* SEKI-Report SR–2009–01 (ISSN 1437–4447). SEKI Publications, DFKI Bremen GmbH~~, Safe and Secure Cognitive Systems, Cartesium, Enrique Schmidt Str. 5, D-28359 Bremen,~~ Germany, 2014. Rev. edn. May 2014, ii+82 pp., http://arxiv.org/abs/0902.4682. [see above]

[WIRTH, 1997] Claus-Peter Wirth. *Positive/Negative-Conditional Equations: A Constructor-Based Framework for Specification and Inductive Theorem Proving*, volume 31 of *Schriftenreihe Forschungsergebnisse zur Informatik*. Verlag Dr. Kovač, Hamburg, 1997. PhD thesis, Univ. Kaiserslautern, ISBN 386064551X, http://wirth.bplaced.net/p/diss.

[WIRTH, 2009] Claus-Peter Wirth. Shallow confluence of conditional term rewriting systems. *J. Symbolic Computation*, 44:69–98, 2009. http://dx.doi.org/10.1016/j.jsc.2008.05.005.

[WIRTH, 2012] Claus-Peter Wirth. HERBRAND's Fundamental Theorem in the eyes of JEAN VAN HEIJENOORT. *Logica Universalis*, 6:485–520, 2012. ~~Received Jan. 12, 2012. Published online June 22,~~ 2012, http://dx.doi.org/10.1007/s11787-012-0056-7. [not needed]

[WIRTH, 2014] Claus-Peter Wirth. HERBRAND*'s Fundamental Theorem: The Historical Facts and their Streamlining.* SEKI-Report SR–2014–01 (ISSN 1437–4447). SEKI Publications, DFKI Bremen GmbH, ~~Safe and Secure Cognitive Systems, Cartesium, Enrique Schmidt Str. 5, D-28359 Bremen,~~ Germany, 2014. ii+47 pp., http://arxiv.org/abs/1405.6317. [delete!]