Information and Computation 255 (2017) 287–310

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Rethinking specificity in defeasible reasoning and its effect in argument reinstatement ☆

Gustavo Adrián Bodanza [a,*], Claudio Andrés Alessio [b]

[a] *Departamento de Humanidades, Universidad Nacional del Sur, and Instituto de Investigaciones Económicas y Sociales del Sur, CONICET, Argentina*
[b] *Universidad Católica de Cuyo, and Instituto de Investigaciones Económicas y Sociales del Sur, CONICET, Argentina*

A R T I C L E I N F O

Article history:
Received 15 February 2015

Keywords:
Argument systems
Specificity
Reinstatement
Defeat
Warrant

A B S T R A C T

The principle of reinstatement governing most argument systems states that an argument is reinstated when all its defeaters are in turn ultimately defeated. Nevertheless, some criticisms to this principle have been offered in the literature. We found that problems arise when arguments in a chain of attacks are related by specificity: when non-maximally specific arguments are reinstated, fallacious justifications are originated. Particularly, we show how the problem affects DeLP, a system that combines a specificity-based defeat criterion with a reinstatement-based warrant process. Following old intuitions by philosopher Carl Hempel we rethink the concept's role within defeasible argumentation. Two kinds of specificity defeaters are identified: proper defeaters and cautious defeaters. While proper defeaters are well-known, cautious defeaters are formally introduced here. A system combining cautious and proper defeaters is defined as an extension of DeLP, and dialectic warrant games are proposed for filtering out non-maximally specific arguments.

1. Introduction

Argument reinstatement is at the core of most argument systems, especially those which can be treated as instances of Dung's argumentation frameworks [11]. The intuition is that an argument should be reinstated when all its possible defeaters are in turn defeated outright. More precisely, A should be warranted if for every defeater B of A there exists some warranted (according to some specific semantics) argument C such that C defeats B (cf. [2]). The example below, introduced by Dung as a motivation for his semantics, illustrates the rationale of reinstating argument A given argument C.

Example 1.

A: (Agent 1:) My government cannot negotiate with your government because your government doesn't even recognize my government.
B: (Agent 2:) Your government doesn't recognize my government either.
C: (Agent 1:) But your government is a terrorist government.

☆ This paper extends and reports further developments of the work in [5].
* Corresponding author.
E-mail address: ccbodanz@criba.edu.ar (G.A. Bodanza).

http://dx.doi.org/10.1016/j.ic.2017.01.006

$$C \longrightarrow B \longrightarrow A$$

Fig. 1. The reinstatement principle: a defeated argument is reinstated if all its defeaters are in turn defeated by some accepted arguments.

Accepting that A and B are mutually attacking arguments and that C attacks B (but not the converse), the reinstatement of A by C makes sense.

On the other hand, other examples suggest that reinstatement cannot be taken as a general principle (Fig. 1):

Example 2.

A: Tweety flies because it is a bird, and birds tend to fly.
B: Tweety does not fly because it is a penguin, and penguins tend not to fly.
C: Tweety flies because it is a magic penguin, and magic penguins tend to fly.

Horty [21] has argued against reinstatement using a similar example. If A and C are jointly admitted, then a sound conclusion (Tweety flies) could be justified on basis of a weak reason (that it flies because it is a bird). Clearly, a stronger reason is that Tweety has a skill that specifically magic penguins have. The acceptance of A would be unsound if the model is intended to offer the best explanation for the conclusion it yields. A would be acceptable only if all the subclasses of birds (including penguins and magic penguins) are equally plausible to fly; but that is not the case here. The fact that C reinstates A's conclusion (which is also C's own conclusion) cannot be a reason for C to reinstate the whole argument A, because A does not meet that criterion.

A worse situation arises when the conclusion of the reinstated argument is stronger than that of the reinstating argument, as in the following case (also introduced by Horty):

Example 3.

A: Beth is millionaire because he is an MS employee, and they tend to be millionaire.
B: Beth has less than half a million because he is a new MS employee, and they tend to have less than half a million.
C: Beth has at least half a million because he is a new MS employee in department X, and they have at least half a million.

Here A's conclusion is stronger than C's conclusion, in the sense that having at least half a million is a logical consequence of being a millionaire but not vice versa. Argument A would be reasonably accepted only if being a millionaire is equally plausible for any subclass of the class of MS employees. Otherwise, the acceptance of A is counterintuitive.

Curiously, all the counterexamples to reinstatement that we found in the literature involve arguments that can be compared by specificity. That motivated the present study, which tries to show that the problem is the way in which specificity is used to establish defeat rather than a problem with the reinstatement principle. In fact, we can find cases where reinstatement is not at stake but a comparison by specificity should be anyway used to reject some too credulous arguments. For instance, imagine a scenario just like that of Example 3 but where argument B is missing.

Example 4.

A: Beth is millionaire because he is an MS employee, and they tend to be millionaire.
C: Beth has at least half a million because he is a new MS employee in department X, and they have at least half a million.

It would be very seldom to find a defeasible reasoning formalism rejecting argument A just because of argument C. Since there is no contradiction between them, C would rarely be considered a defeater of A. Nevertheless, C is more specific than A and its conclusion is more cautious. Clearly, the problem we have seen in Example 3 persists here. In a more recent work, Horty [22]—in agreement with moral philosopher Jonathan Dancy – endorses the "general idea that certain defeaters – particularly those whose priority is derived from specificity – exclude as well as defeat" and claims that this idea "finds further support from cases in which specificity seems to support exclusion even without ordinary defeat" (pp. 325–326). He offers another example:

Example 5. Assume that all least ruffed finches are ruffed finches.

A: Frank is a ruffed finch, and ruffed finches' nests are largely confined to Green Island, hence we can tentatively conclude that Frank's nest is in Green Island.
B: Frank is a least ruffed finch, and least ruffed finches' nests are distributed almost evenly between Green Island and Sand Island, hence we only can conclude tentatively that Frank's nest is either in Green Island or in Sand Island.

The evidence that Frank is a least ruffed finch is more specific than the evidence that Frank is a ruffed finch, so we only want to conclude that he lives in one of the two islands. Again, the more specific argument is more cautious as it supports a weaker conclusion than that of the less specific argument. Not only the conclusion that Frank nests in Green Island should be excluded but also the reasons for that conclusion. As argument B is accepted, argument A is rejected. Unfortunately, current defeasible reasoning systems that formalize a specificity criterion are not able to handle the information in this way.

1.1. Proposal and organization of the paper

What, then, is the point about reinstatement? Our argument is that reinstatement is right in principle and the counterexamples offered in the literature are wrong about it. What is wrong about the counterexamples is that some key defeats are disregarded while others are misconceived, resulting in counterexamples to the way the specificity criterion is implemented rather than to the reinstatement principle. We argue that what happens in cases like Example 3, for instance, is that C defeats A in some way that has to do with specificity (even without contradiction). If this defeat is accepted, then C cannot be considered to reinstate A. On the other hand, we also argue that the defeat by specificity among conflicting arguments – like those from B to A and from C to B – is a strong kind of defeat, in the sense that the defeated arguments result defeated outright and, then, not liable to reinstatement. Both kinds of defeat show that the defeated arguments are not maximally specific: considering the total evidence, such arguments are untenable and hence – we claim – they cannot be reinstated.

This does not mean that reinstatement is invalid in general. If other kinds of defeat can be defined among maximally specific arguments, then reinstatement is right. In Example 1, for instance, all the arguments are (ceteris paribus) maximally specific and reinstatement yields the expected results. The fact that the attacks among them are not established on a comparison by specificity is not just a coincidence.

Prakken [30] argued that reinstatement cannot be applied when statistical reasoning is at stake because more general arguments just cannot be constructed in a right representation, since the pertinent defaults must be blocked so that only the most specific arguments remain; hence – Prakken concludes – the problem here is not about reinstatement but one of representation. We agree that the issue of representation cannot be eluded, but while finding general principles of representation could be a hard enterprise, the problem can instead be approached by looking for general conditions under which the arguments can compete and defeat among them. Accordingly, we will argue that a sound redefinition of the concept of specificity can help to state those conditions.

In a preceding work [5], we have identified a condition under which an argument A is undermined, to wit, some evidence on which A is built also enables the construction of some more specific argument B that contradicts a claim asserted in A. Moreover, we have argued that every argument constructed using that evidence can be considered a defeater of A. We called these defeaters *undermining defeaters*. But we have now found that often the evidence undermines an argument even when no contradiction occurs. Example 5 is a case. The evidence that Frank is a least ruffed finch is a reason not to treat Frank as an ordinary ruffed finch to conclude that it nests in Green Island, since least ruffed finches – specifically – tend to nest – more generally – either in Green Island or in Sand Island. This more general conclusion does not contradict the other one, but is more cautious and could prevent a reckless inference. Hence, arguments of this kind should be considered defeaters in some way, even in absence of contradiction.

To give our proposal the due in-depth treatment we analyze early discussions about specificity in philosophy of science. Hempel [19] defended a *requirement of maximal specificity* as a condition for the acceptance of probabilistic or inductive-statistical explanations. Early applications of specificity in non-monotonic reasoning in AI were also aware of the intuition that only maximally specific explanations should be accepted, so from the argumentative point of view [29] as from the defeasible inheritance networks point of view [14,15,23,32]. Nevertheless, all of these approaches suffer some of the mentioned problems.

This analysis has led us to improve the notion of maximal specificity. We identify here some necessary conditions for an argument to be maximally specific, conditions that warn about the possible existence of basically two reasons for rejection. One is the existence of "proper" defeaters, i.e. conflicting arguments based on more evidence, and the other one is the existence of "cautious" defeaters, i.e. arguments showing that the total evidence only enables a not so specific conclusion. While "proper" defeaters are well known ([29,33,18], etc.), the notion of "cautious" defeater is formally introduced here. Contrarily to the notion of undermining defeater introduced in [5], cautious defeaters are defined independently from proper defeaters.

For our purpose we need a formalism capable of expressing evidence, defeasible knowledge, strict knowledge, and building and comparing arguments. We think that DeLP's language [18] – an evolved version of Simari and Loui's model [33] – is the most suitable formalism as it was performed to compare arguments by specificity as well as to justify arguments by a warrant procedure satisfying reinstatement.

Since we claim that non-maximally specific arguments should be rejected outright (thus, unable to be reinstated), we define a warrant criterion that filters them out. It takes the form of a dialectic argumentation game between a proponent and an opponent, where players advance arguments in turns, each attacking the last move of the other player. In the process, arguments can be defeated and reinstated until some player is out of moves. Alternative protocols for skeptic and credulous behaviors are also defined, both of them containing rules that penalize the use of non-maximally specific arguments. In each case, the winning strategies of the proponent determine the warranted arguments.

The paper is organized as follows. In section 2, the DeLP system [18] is briefly outlined. Some of the above mentioned examples are modeled with that formalism, showing how counterintuitive behaviors are obtained. In section 3, the ideas of philosopher Carl Hempel about the specificity of inductive explanations are analyzed. That will help us to define a criterion for defeasible argumentation. The arising notion of cautious defeater is introduced next in subsection 3.3, where some formal properties are also shown. A procedure for finding them is outlined in 3.4. In section 4, the dialectical game for warrant is proposed. Further examples and related opinions are discussed in section 5. Finally, we sum up our findings is section 6.

2. The DeLP formalism

The DeLP formalism [18] is based on a first-order language $\mathcal{L}$ that is partitioned in three disjoint sets: a set of facts, a set of strict rules and a set of defeasible rules. *Facts* are literals, i.e. ground atoms (L) or negated ground atoms ($\sim L$, where '$\sim$' represents the classical negation); facts represent particular knowledge. Both *strict* and *defeasible rules* are program rules. Syntactically, strict rules are sequents of the form $L \leftarrow L_1, \ldots, L_n$ and defeasible rules are sequents of the form $L \prec L_1, \ldots, L_n$, where $L, L_1, \ldots, L_n$ are literals. Strict rules represent general, non-defeasible knowledge while defeasible rules represent tentative, defeasible knowledge. A *defeasible logic program* (de.l.p.) $\mathcal{P}$ is a pair (Π, Δ) where Π is a consistent set of formulae partitioned in two subsets Π_F, containing only facts, and Π_G, containing only strict rules, and Δ is a set of defeasible rules. Given a de.l.p. $\mathcal{P} = (\Pi, \Delta)$ we say that a literal L is a *defeasible derivation* from Γ in $\mathcal{P}$, in symbols, $\Gamma \mid\!\sim_{\mathcal{P}} L$ iff $\Gamma \subseteq \Pi \cup \Delta$ and there exists a sequence of ground (instantiated) literals $L_1, \ldots, L_n$ such that $L_n = L$ and for each L_i, $1 \leq i \leq n$, either $L_i \in \Gamma$ or there exists either a strict rule $(L \leftarrow L'_1, \ldots, L'_k)$ or a defeasible rule $(L \prec L'_1, \ldots, L'_k)$ in Γ such that $\{L'_1, \ldots, L'_k\} \subseteq \{L_1, \ldots, L_{i-1}\}$. If all the rules used in the derivation of L are strict then we say that L is a *strict derivation* from Γ, in symbols, $\Gamma \vdash_{\mathcal{P}} L$. (From now on, we will write '$\mid\!\sim$' and '$\vdash$' instead of '$\mid\!\sim_{\mathcal{P}}$' and '$\vdash_{\mathcal{P}}$', respectively, when the referenced de.l.p. is obvious.)

Definition 1 *(Argument structure [18]).* Given a de.l.p. $\mathcal{P} = (\Pi, \Delta)$, an *argument structure* (in $\mathcal{P}$) is a pair $\langle T, h \rangle$, where $T \subseteq \Delta$ and h is a literal (the argument's *conclusion*), and

1. $\Pi \cup T \mid\!\sim h$,
2. $\Pi \cup T \not\mid\!\sim \bot$,
3. $\nexists T'(T' \subset T \ \wedge \Pi \cup T' \mid\!\sim h)$.

That is, an argument structure is a subset of defeasible rules that, together with the strict knowledge, enables the defeasible derivation of the argument's conclusion, does not lead to the defeasible derivation of contradictions and is minimal, i.e. does not contain unnecessary rules for the derivation of the conclusion. This minimality condition is fundamental for the definition of the specificity relation in DeLP.

2.1. Specificity in DeLP

Generalized specificity is one of the two comparison criteria among arguments which is especially defined in DeLP,[1] which was formerly adopted by Simari and Loui [33] from Poole [29]. Poole, in time, proposes a semantical formulation of specificity to solve ambiguity problems which arise in inheritance systems [20]. He defines specificity as a comparison criterion among *explanations*, *theories* or *solutions*, which are mutatis mutandis the same as what in DeLP are called *arguments*. In DeLP, as in [33], the following syntactic version is offered – though it can be seen just as a paraphrasis of Poole's semantic formulation.

Definition 2 *(Strictly more specific [18]).* Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. and let F be the set of all literals that have a defeasible derivation from $\mathcal{P}$. Let $\langle T_1, h_1 \rangle$ and $\langle T_2, h_2 \rangle$ be two argument structures obtained from $\mathcal{P}$. $\langle T_1, h_1 \rangle$ is *strictly more specific than* $\langle T_2, h_2 \rangle$, in symbols, $\langle T_1, h_1 \rangle \succ_{\mathbf{spec}} \langle T_2, h_2 \rangle$ iff

1. for all $H \subseteq F$, if $H \cup \Pi_G \cup T_1 \mid\!\sim h_1$ and $H \cup \Pi_G \nvdash h_1$ then $H \cup \Pi_G \cup T_2 \mid\!\sim h_2$ (every H that "activates" T_1 also "activates" T_2), and
2. there exists $H \subseteq F$ such that $H \cup \Pi_G \cup T_2 \mid\!\sim h_2$, $H \cup \Pi_G \nvdash h_2$ and $H \cup \Pi_G \cup T_1 \not\mid\!\sim h_1$ (some H "activates" T_2 but not T_1).

The term 'activates' refers to the relation between a set of literals H and an argument structure $\langle T, h \rangle$ such that, together with Π_G, H and T are enough to construct a (nontrivial) defeasible derivation of the conclusion h. In sum, an argument structure $\langle T_1, h_1 \rangle$ is strictly more specific than another argument structure $\langle T_2, h_2 \rangle$ if every possible evidence which makes T_1 explain h_1 also makes T_2 explain h_2, but some possible evidence makes T_2 explain h_2 while it does not make T_1 explain h_1. Let us see now the behavior of this specificity criterion with respect to some of the previous examples.

[1] The other one is rule's priorities.

Example 6 *(Example 2 formalized).* Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. representing the knowledge that all magic penguins are penguins, all penguins are birds, birds tend to fly, penguins tend not to fly, magic penguins tend to fly, and Tweety is a magic penguin:

$$\Pi_G = \{bird(x) \leftarrow penguin(x), penguin(x) \leftarrow magic\_penguin(x)\}$$

$$\Pi_F = \{magic\_penguin(Tweety)\}$$

$$\Delta = \{flies(x) \prec bird(x), \sim flies(x) \prec penguin(x),$$
$$flies(x) \prec magic\_penguin(x)\}$$

Then we have the argument structures:

$$A = \langle \{flies(Tweety) \prec bird(Tweety)\}, flies(Tweety)\rangle,$$

$$B = \langle \{\sim flies(Tweety) \prec penguin(Tweety)\}, \sim flies(Tweety)\rangle,$$

$$C = \langle \{flies(Tweety) \prec magic\_penguin(Tweety)\}, flies(Tweety)\rangle.$$

Then $C \succ_{\mathbf{spec}} B$, $C \succ_{\mathbf{spec}} A$, and $B \succ_{\mathbf{spec}} A$.

This example could suggest that all we need is to pick just the most specific argument in terms of $\succ_{\mathbf{spec}}$ (i.e. its maximal element) and discard the rest. Indeed, that criterion would also solve Examples 3 and 5, and even Example 4. But though this idea could be right for usual cases it can have the effect of precluding acceptable arguments in other cases where less specific arguments are not in conflict with the more specific one. That is what would happen in the following example:

Example 7. Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. representing the knowledge that all lapwings are birds, birds tend to fly, lapwings tend to nest on the ground and Pedro is a lapwing:

$$\Pi = \{bird(x) \leftarrow lapwing(x), lapwing(Pedro)\}$$

$$\Delta = \{flies(x) \prec bird(x), nests\_on\_the\_ground(x) \prec lapwing(x)\}$$

Then we have the argument structures:

$$A = \langle \{flies(Pedro) \prec bird(Pedro)\}, flies(Pedro)\rangle,$$

$$B = \langle \{nests\_on\_the\_ground(Pedro) \prec lapwing(Pedro)\},$$
$$nests\_on\_the\_ground(Pedro)\rangle$$

Since $B \succ_{\mathbf{spec}} A$, by choosing only the maximal elements of $\succ_{\mathbf{spec}}$ the acceptable argument A and its conclusion $flies(Pedro)$ are both precluded.

Poole [29] proposes to prefer the most specific explanation not only in cases of ambiguity or contradiction. Simari and Loui [33] and García and Simari [18], instead, use specificity as the comparison criterion involved in a defeat relation for which contradiction is an indispensable trigger: an argument A defeats an argument B if there exists some subargument B' of B such that A is in contradiction with, and is strictly more specific than, B'.

Definition 3 *(Subargument [18]).* An argument structure $\langle T', h'\rangle$ is a *subargument structure* of an argument structure $\langle T, h\rangle$ iff $T' \subseteq T$.

Definition 4 *(Proper defeater [18]).* An argument structure $\langle S, j\rangle$ is a *proper defeater* of an argument structure $\langle T, h\rangle$ if for some sub-argument $\langle T', h'\rangle$ of $\langle T, h\rangle$, $\langle S, j\rangle \succ_{\mathbf{spec}} \langle T', h'\rangle$ and $\Pi \cup \{j, h'\} \vdash \perp$. Given a set of argument structures S we also define $def_{prop}(S) =_{df} \{(A, B) : A, B \in S$ and A is a proper defeater of $B\}$.

Unfortunately, this approach is insufficient to solve the problem. Consider Example 6 again. Though $C \succ_{\mathbf{spec}} A$, C is not a proper defeater of A because there is no contradiction between them, and so argument A results undefeated. That is, not only the conclusion *flies*(*Tweety*) is reinstated – which is right in this case – but also the explanation for that, which is that Tweety is a bird and birds tend to fly, is accepted. In Example 3 we have a worse problem, since a wrong conclusion comes out reinstated which is supported by an unacceptable reason:

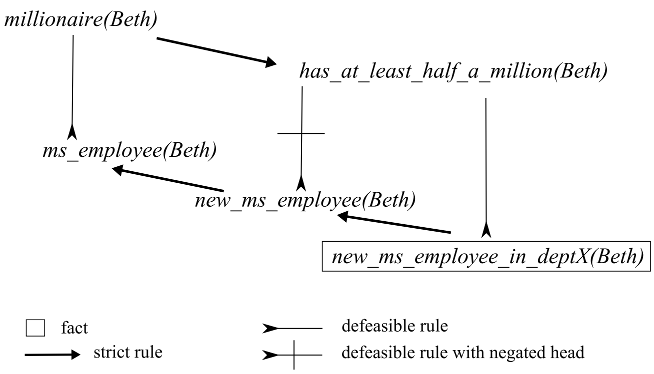

Fig. 2. De.l.p. of Example 8.

Example 8 *(Example 3 formalized).* Let $\mathcal{P} = (\Pi, \Delta)$ a de.l.p. such that

$$\begin{aligned}\Pi = \{\, & has\_at\_least\_half\_a\_million(x) \leftarrow millionaire(x),\\ & ms\_employee(x) \leftarrow new\_ms\_employee(x),\\ & new\_ms\_employee(x) \leftarrow new\_ms\_employee\_deptX(x),\\ & new\_ms\_employee\_deptX(Beth)\}\\ \Delta = \{\, & millionaire(x) \prec ms\_employee(x),\\ & \sim has\_at\_least\_half\_a\_million(x) \prec new\_ms\_employee(x),\\ & has\_at\_least\_half\_a\_million(x) \prec new\_ms\_employee\_deptX(x)\}\end{aligned}$$

Then we have the argument structures:

$$\begin{aligned}A &= \langle\{millionaire(Beth) \prec ms\_employee(Beth)\}, millionaire(Beth)\rangle,\\ B &= \langle\{\sim has\_at\_least\_half\_a\_million(Beth) \prec new\_ms\_employee(Beth)\},\\ &\quad \sim has\_at\_least\_half\_a\_million(Beth)\rangle,\\ C &= \langle\{has\_at\_least\_half\_a\_million(Beth) \prec new\_ms\_employee\_deptX(Beth)\},\\ &\quad has\_at\_least\_half\_a\_million(Beth)\rangle.\end{aligned}$$

Then C is a proper defeater of B and B is a proper defeater of A. And though $C \succ_{\mathbf{spec}} A$, C is not a proper defeater of A. Hence, A is accepted, in agreement with the reinstatement principle (Fig. 2).

3. Rethinking specificity

3.1. Making specificity a transitive relation

An important problem in Poole's original formulation of the specificity relation is the lack of transitivity. Recently, Wirth and Stolzenburg [37] have discussed this shortcoming, which certainly is inherited by the formulation in DeLP (for details about the issue we refer the reader to the mentioned work, especially to Example 6, which shows a counterexample of transitivity). After analyzing the source of the problem, the authors propose a new relation of specificity, called 'CP1', which we adopt here. The difference lies in the *activation sets* taken into account to define each specificity relation: while Poole takes as activation sets the subsets of literals occurring in formulae of either Π or Δ, Wirth and Stolzenburg take subsets of literals occurring only in formulae of Π. For a formal presentation, we adapt these ideas to DeLP's formalism as follows.

Definition 5 *(Theory of a program [37]).* Given a program $\mathcal{P} = (\Pi, \Delta)$, the *theory* of $\mathcal{P}$ is the set $\mathcal{T}_{\mathcal{P}}$ of all the literals L such that either 1) an instance of L is in $\mathcal{P}$, or 2) there is a conjunction C of literals from $\mathcal{T}_{\mathcal{P}}$ such that either $L \prec C$ or $L \leftarrow C$ is an instance of a rule in $\mathcal{P}$.

Definition 6 *(Activation set [37]).* H is an *activation set* for an argument $\langle T, h\rangle$ if $h \in \mathcal{T}_{\mathcal{L}\cup\Pi}$ for some $\mathcal{L} \subseteq \mathcal{T}_{H\cup T\cup\Pi_G}$.

Definition 7 *($CP1$ (at least as specific as) [37]).* Let $\langle T_1, h_1\rangle$ and $\langle T_2, h_2\rangle$ be two argument structures. Then $\langle T_1, h_1\rangle \leq_{CP1} \langle T_2, h_2\rangle$[2] iff either

1. $h_1 \in \mathcal{T}_\Pi$, or
2. $h_2 \notin \mathcal{T}_\Pi$ and every $H \subseteq \mathcal{T}_\Pi$ that is an activation set for $\langle T_1, h_1\rangle$ is also an activation set for $\langle T_2, h_2\rangle$.

Theorem 1. *$\leq_{CP1}$ is a quasi-ordering relation on arguments (i.e. reflexive and transitive).*

Proof. See [37], Theorem 2.

The corresponding equivalence and strict order are obtained as usual:

$$A \approx_{CP1} B =_{df} A \leq_{CP1} B \wedge B \leq_{CP1} A$$
$$A <_{CP1} B =_{df} A \leq_{CP1} B \wedge B \nleq_{CP1} A$$

Though in general it does not hold that $\succeq_{\mathbf{spec}} \subseteq \leq_{CP1}$, Wirth and Stolzenburg [37] provide sufficient conditions under which it holds (Theorem 3). In particular, it holds if all the rules, strict or defeasible, have no more than one literal in their antecedents. This covers all our examples so far. On the other hand, $\leq_{CP1}$ improves the specificity relation with respect to $\succeq_{\mathbf{spec}}$ as a preference relation, but the problems observed before are still not solved. One interesting improvement is that transitivity guarantees the non-existence of cycles in the relation. On the other hand, transitivity could suggest that choosing only the maximal elements of $\leq_{CP1}$ is enough to get the "best"competing arguments. But, as we have seen in Example 7, this policy could lead to exclude some arguments that should be kept. That is, non-maximal elements of a set of arguments ordered w.r.t. $\leq_{CP1}$ are not always "bad" or "worse" than the maximal ones. In that example, both arguments (i.e. Pedro nests on the ground because it is a lapwing, and Pedro flies because it is a bird) are maximally specific in some sense, one that cannot be captured only by $\leq_{CP1}$ (and, certainly, neither by $\succeq_{\mathbf{spec}}$). That sense has to do with how arguments are logically related. In part, a maximally specific argument is such that there exists no other contradicting argument which uses more specific evidence (that is, one that has no proper defeaters). But we have identified other conditions as well. This is the subject of the next subsection, where we bring back to discussion old ideas from Philosophy of Science to explore a wider sense of "maximally specific".

3.2. On maximal specificity

Carl Hempel [19] proposed a requirement of maximal specificity as an acceptability criterion for probabilistic or inductive-statistical explanations (see [17] for a brief review). The intuition is that what is inferred in a maximally specific explanation about a class G taking into account the total evidence must also be inferred about any subclass H of G with the same probability. Given a knowledge base $\mathcal{K}$ (which includes the total evidence, natural laws, axioms of the probability theory, etc.) a probabilistic explanation of a fact Gi has the form

$$\begin{array}{ll} Fi & \\ \underline{\underline{p(G|F) = r}} & \\ Gi & [r] \end{array}$$

where $p(G|F)$ is the probability of an F to be a G, r is a number between 0 and 1 which is sufficiently close to 1, Fi expresses the fact that i is an F, and $[r]$ is the inductive probability of the *explanandum* Gi relative to the *explanans* $\{(p(G|F) = r, Fi\}$. Then, Hempel argues that this is a maximally specific explanation only if for any class F' for which $\mathcal{K}$ contains statements to the effect that F' is a subclass of F and that $F'i$, $\mathcal{K}$ also contains a probabilistic-statistical law to the effect that $p(G|F') = r'$, where $r' = r$ unless the law is a theorem of probability theory.[3]

Clearly, a simple extrapolation of this criterion of maximal specificity to defeasible argumentation is not suitable, since defeasible inferences have no probability measures. Nevertheless, we can discern the same underlying intuition about the maximal specificity of both probabilistic and defeasible inferences: more specific evidence cannot diminish the strength of the inference. This leads us to think about a qualitative counterpart of probability in terms of defeasibility. Next we argue on how this can be approached, by providing two necessary conditions for a defeasible argument to be maximally specific.

Firstly, let us consider that, from Hempel's requirement, we know that if $\{p(G|F) = r, Fi\}$ is a maximally specific explanation of Gi, then there exists no subclass F' of F such that $p(\sim G|F') \geq r$. In terms of defeat, that means that $p(\sim G|F') \geq r$ is a reason to defeat the explanation of Gi. Hence, a maximal specific defeasible argument should be such that the total

---

2 Note that the direction of the symbol '$\leq_{CP1}$' is reversed with respect to '$\succ_{\mathbf{spec}}$'.

3 "The 'unless'-clause is meant to allow $\mathcal{K}$ to contain [...] pairs of sentences such as '$Fi \wedge Gi$' and '$p(G|F \wedge G) = 1$'; the latter, being a theorem of the probability calculus, is thus not reckoned as an explanatory empirical law" (Hempel, [19]: 120).

evidence[4] does not enable the derivation of the contrary conclusion. In our setting, this means that a maximally specific argument is one with no proper defeaters. To express this condition we use the transitive relation $<_{CP1}$ instead of $\succ_{\mathbf{spec}}$.[5] Moreover, we have to take into account the compositional aspect of an argument: the requirement applies to all the subarguments which compose an argument. Hence, we formulate the condition as follows:

(C1) *An argument $\langle T, h\rangle$ is maximally specific only if for every subargument $\langle T', h'\rangle$ of $\langle T, h\rangle$ and every argument $\langle S, j\rangle$ such that $\langle S, j\rangle \leq_{CP1} \langle T', h'\rangle$, $\Pi \cup \{j, h'\} \nvdash \perp$.*

Clearly, this condition prevents from the existence of a proper defeater. That is, the mere existence of a proper defeater is sufficient for the defeated argument not to be maximally specific, as the evidence that enables its construction also enables the construction of that defeater. Note that it does not matter if the evidence enables, in time, the construction of a defeater of that proper defeater. In this sense, the evidence at hand is sufficient to undermine the argument inasmuch it diminishes the plausibility of the conclusion. In Example 6, for instance, argument C satisfies (C1), while neither B nor A satisfy it: A does not because B is a proper defeater, so the evidence that Tweety is a penguin weakens the argument that Tweety flies because it is a bird; and B is not maximally specific neither because C is a proper defeater of it; the evidence that Tweety is a magic penguin weakens the argument that Tweety does not fly because it is a penguin. Note also that (C1) can be used to prevent the reinstatement of A: magic penguins are a proper subclass of penguins, which in time are a proper subclass of birds, and the property of flying is not inherited through those nested classes with the same plausibility (note the analogy with Hempel's criterion above). The case of Example 8 is similar.

Secondly, we have found in situations like that of Example 3 that specific information can determine, in Stephen Toulmin's words, other "circumstances in which the general authority of the warrant would have to be set aside" [35]: 94). Consider again Example 4 – the formal expression of Example 3. The defeasible rule

$$has\_at\_least\_half\_a\_million(x) \prec new\_ms\_employee\_deptX(x)\}$$

together with the evidence that Beth is a new employee in department X, makes plausible the conclusion that Beth has at least half a million. Why should we represent the information contained in this rule? The answer, of course, is that for cases like Beth we do not want to conclude that they are millionaires, we just want to conclude they have at least half a million. Note that we do not want to conclude that they are not millionaires neither. If the program is used, for instance, for a tentative stratification of MS employees regarding their incomes, then the conclusions 'Beth is a millionaire' and 'Beth is not a millionaire' seem both not only inaccurate but less than plausible in the information context. The mentioned defeasible rule is needed to prevent those conclusions: it poses a reason for concluding something weaker than being a millionaire without concluding something so strong as not being a millionaire. In that sense, that rule enables the construction of defeating reasons that, unlike proper defeaters, do not support the opposite conclusion.

Our argument is clearly pragmatic, since it is based on representational needs. To reinforce this argument, let us imagine a probabilistic version of Example 4 from which the defeasible reasoning version can be obtained. This can be done by expressing each relevant (and non-redundant) statistical piece of information $p(G|F) = r$ ($r > .5$) as a defeasible rule '$G \prec F$'.[6] We want to show that the explanation for 'Beth is a millionaire' does not qualify as maximally specific under Hempel's requirement in such probabilistic scenario. The reason is that

$$p(millionaire|ms\_employee)$$

should be unequal to

$$p(millionaire|new\_ms\_employee\_deptX).$$

Otherwise, if being a millionaire is equally probable for MS employees in general as for new MS employees in department X in particular, then the defeasible rule saying that these last tend to have half a million would be spurious, since being an MS employee is deducible from being a new MS employee in department X, as having half a million is deducible from being a millionaire. Hence, given those unequal probabilities, the statistical explanation that Beth is a millionaire given that she is an MS employee would not be maximally specific, since being a member of the subclass of new employees in department X confers a different probability to that conclusion.[7] In order to generalize the criterion, we would say that the explanation for E (e.g. Beth is a millionaire) would be maximally specific only if, for every E' which can be explained with the evidence at hand (e.g. Beth has at least half a million), if E' can be explained with the same reasons that explain E and with the same probability, then E can be explained with the same reasons that explain E' and with the same probability. In our example, though, the explanation for 'Beth is a millionaire' does not satisfy this requirement. While 'Beth has at least half a

[4] In our setting, 'total evidence' can just be interpreted as formulae in Π (facts representing factual evidence and strict rules representing natural laws).

[5] In order to be precise, we would change the term 'proper defeater' to 'CP1-proper defeater' to name the defeat relation resulting by changing '$\succ_{\mathbf{spec}}$' to '$<_{CP1}$' in Definition 4, and similarly for any other specificity relation considered. Nevertheless, we will avoid any prefix for the sake of simplicity when the context makes clear the meaning.

[6] We are considering here Pollock's ideas on the 'statistical syllogism' (SS): "$p(G|F) = r(r > .5) \wedge Fa$" is a defeasible reason for "Ga" ([28]: 235).

[7] Cf. also Pollock's *subproperty defeaters* as expressing a total evidence requirement: "$Ha \wedge p(G|F \wedge H) \neq p(G|F)$" is an undercutting defeater for (SS) ([28]).

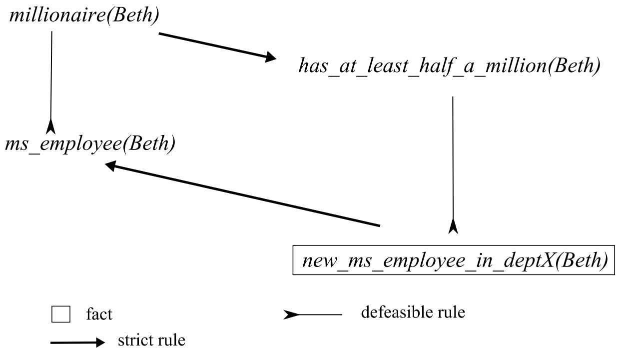

Fig. 3. De.l.p. of Example 9.

million' can be explained with the same reasons that explain 'Beth is a millionaire' and with the same probability – since being a new employee in department X logically implies being an MS employee, and being a millionaire logically implies having at least half a million–, 'Beth is a millionaire' cannot be explained with the same reasons that explain 'Beth has at least half a million' with the same probability – since being a millionaire is not equally probable for new employees in department X. Analogously, the general criterion for a maximally specific defeasible argument can be stated as follows (again, the compositional aspect is considered):

(C2) *An argument $\langle T, h\rangle$ is maximally specific only if for every subargument $\langle T', h'\rangle$ of $\langle T, h\rangle$ and every argument $\langle S, j\rangle$ such that $\langle S, j\rangle \leq_{CP1} \langle T', h'\rangle$, if $\langle T', j\rangle$ is an argument structure then $\langle S, h'\rangle$ is also an argument structure.*

That is, if the more general argument (T') can explain the same conclusion explained by the more specific argument (S), then the more specific argument can explain the same conclusion as the more general one. Otherwise, the more specific argument supports a possible exception to the reasons of the more general one.

Example 9 *(Example 4 revisited).* The argument

$$A = \langle\{millionaire(Beth) \multimap ms\_employee(Beth)\}, millionaire(Beth)\rangle$$

does not satisfy (C2) since there exists a fact, to wit,

$$F = \{new\_ms\_employee\_deptX(Beth)\},\ \text{which activates the argument}$$

$$C = \langle\{has\_at\_least\_half\_a\_million(Beth) \multimap new\_ms\_employee\_deptX(Beth)\},$$
$$has\_at\_least\_half\_a\_million(Beth)\rangle,$$

and while we can construct the argument

$$\langle\{millionaire(Beth) \multimap ms\_employee(Beth)\},$$
$$has\_at\_least\_half\_a\_million(Beth)\rangle$$

we cannot construct the argument

$$\langle\{has\_at\_least\_half\_a\_million(Beth) \multimap new\_ms\_employee\_deptX(Beth)\}, millionaire(Beth)\rangle \text{ (Fig. 3).}$$

Example 10 *(Example 7 revisited).* Argument

$$B = \langle\{\ nests\_on\_the\_ground(Pedro) \multimap lapwing(Pedro)\},$$
$$nests\_on\_the\_ground(Pedro)\rangle$$

satisfies (C2) since no argument X besides B itself is such that $X \leq_{CP1} B$. On the other hand, argument $A = \langle\{flies(Pedro) \multimap bird(Pedro)\}, flies(Pedro)\rangle$, also satisfies the condition, since B is the only argument besides A such that $B \leq_{CP1} A$, but $\langle\{flies(Pedro) \multimap bird(Pedro)\}, nests\_on\_the\_ground(Pedro)\rangle$ is not an argument structure.

We are now in conditions to formalize our notion of maximal specificity for defeasible arguments, combining the requirements (C1) and (C2). Since we are aware that other requirements could be needed for a complete characterization, the conditions are stated only as necessary (not sufficient).

Definition 8 *(Maximal specificity).* Given a de.l.p. $\mathcal{P} = (\Pi, \Delta)$, an argument structure $\langle T, h\rangle$ (in $\mathcal{P}$) is *maximally specific* only if for every subargument $\langle T', h'\rangle$ of $\langle T, h\rangle$ and for every argument structure $\langle S, j\rangle$ such that $\langle S, j\rangle \leq_{CP1} \langle T', h'\rangle$, either $\langle S, h'\rangle$ is an argument structure or

1. $\Pi \cup \{j, h'\} \nvdash \perp$, and
2. $\langle T', j\rangle$ is not an argument structure.

In sum, the requirement says that $\langle T, h\rangle$ is maximally specific only if, for every argument based on more specific evidence which does not explain the same conclusion, it explains neither the contrary nor something more general. In other words, a maximally specific argument is one such that the conclusion is "normal" for the class of things the argument talks about, and there exists no reason for exception considering the total evidence.

3.3. Cautious defeaters

The second condition we stated for an argument to be maximally specific (C2) requires cautiousness. The notion of 'cautiousness' can be precised as a relation among argument structures as follows:

Definition 9 *(Strictly more cautious than).* Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. and let $\langle T_1, h_1\rangle$ and $\langle T_2, h_2\rangle$ be two argument structures obtained from $\mathcal{P}$. $\langle T_1, h_1\rangle$ is *strictly more cautious than* $\langle T_2, h_2\rangle$, in symbols, $\langle T_1, h_1\rangle \succ_{\mathbf{cau}} \langle T_2, h_2\rangle$ iff

1. $\langle T_1, h_1\rangle <_{CP1} \langle T_2, h_2\rangle$,
2. $\langle T_2, h_1\rangle$ is an argument structure (in $\mathcal{P}$),
3. $\langle T_1, h_2\rangle$ is not an argument structure (in $\mathcal{P}$).

Roughly, we are saying that $\langle T_1, h_1\rangle$ is more cautious than $\langle T_2, h_2\rangle$ if it uses more specific evidence to explain something that can also be explained by the other argument using more general evidence, while that explained by the more general argument cannot be explained by the more specific one. Cautiousness is about that specific evidence which makes h_1 more plausible than h_2.

Example 11 *(Example 8 revisited).* Argument C is strictly more cautious than argument A. Note that the activation sets of A are the elements of

$$Pow(\{\, ms\_employee(Beth),$$
$$new\_ms\_employee(Beth),$$
$$new\_ms\_employee\_deptX(Beth)\}) \setminus \{\emptyset\}$$

while the only activation set of C is $\{new\_ms\_employee\_deptX(Beth)\}$, hence $C \leq_{CP1} A$ (but $A \nleq_{CP1} C$).[8] Moreover,

$$\langle \{\, millionaire(Beth) \prec ms\_employee(Beth)\},$$
$$has\_at\_least\_half\_a\_million(Beth)\rangle$$

is an argument structure but

$$\langle \{\, has\_at\_least\_half\_a\_million(Beth) \prec new\_ms\_employee\_deptX(Beth)\},$$
$$millionaire(Beth)\rangle$$

is not an argument structure. On the other hand, neither C and B nor A and B are related by $\succ_{\mathbf{cau}}$.

It is also important to define the conditions under which two argument structures are equally cautious.

Definition 10 *(Equally cautious).* Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. and let $\langle T_1, h_1\rangle$ and $\langle T_2, h_2\rangle$ be two argument structures obtained from $\mathcal{P}$. $\langle T_1, h_1\rangle$ and $\langle T_2, h_2\rangle$ are *equally cautious*, in symbols, $\langle T_1, h_1\rangle \approx_{\mathbf{cau}} \langle T_2, h_2\rangle$ iff

1. $\langle T_1, h_1\rangle \approx_{CP1} \langle T_2, h_2\rangle$,
2. $\langle T_2, h_1\rangle$ is an argument structure (in $\mathcal{P}$),
3. $\langle T_1, h_2\rangle$ is an argument structure (in $\mathcal{P}$).

---

[8] $Pow(S)$ is the power set of the set S.

Lemma 1. *The relation $\approx_{\mathbf{cau}}$ is an equivalence relation, i.e. reflexive, symmetric and transitive.*

Proof. Immediate from definition. □

The concept of 'at least as cautious as' is obtained by the combination of the above two notions.

Definition 11 *(At least as cautious as).* We say that $\langle T_1, h_1\rangle$ is *at least as cautious as* $\langle T_2, h_2\rangle$, in symbols, $\langle T_1, h_1\rangle \succeq_{\mathbf{cau}} \langle T_2, h_2\rangle$ iff either $\langle T_1, h_1\rangle \succ_{\mathbf{cau}} \langle T_2, h_2\rangle$ or $\langle T_1, h_1\rangle \approx_{\mathbf{cau}} \langle T_2, h_2\rangle$.

The following lemma helps to reduce the search space for the most cautious arguments.

Lemma 2. *Let $\langle T_1, h_1\rangle$ and $\langle T_2, h_2\rangle$ be two argument structures in a program $\mathcal{P} = (\Pi, \Delta)$. If $\Pi \cup \{h_2\} \vdash h_1$ then $\langle T_2, h_1\rangle$ is an argument structure (in $\mathcal{P}$).*

Proof. It follows immediately from the definition of *argument structure*. □

The above result ensures that for any argument $\langle T_1, h_1\rangle$, if an argument $\langle T_2, h_2\rangle$ is found such that $\langle T_2, h_2\rangle <_{CP1} \langle T_1, h_1\rangle$, then to see if $\langle T_2, h_2\rangle$ is more cautious than $\langle T_1, h_1\rangle$ it suffices to find any sequence of strict rules $h_1 \to \ldots \to h_2$ and to check that no sequence $h_2 \to \ldots \to h_1$ exists.

Theorem 2. *The relation $\succeq_{\mathbf{cau}}$ is reflexive and transitive.*

Proof. It is obvious that $\succeq_{\mathbf{cau}}$ is reflexive, since for every argument A, $A \le_{CP1} A$ and the conclusion of A is a logical consequence of itself. Transitive: We only have to prove that $\succ_{\mathbf{cau}}$ is transitive. Assume $A \succ_{\mathbf{cau}} B$ and $B \succ_{\mathbf{cau}} C$ and let us prove that $A \succ_{\mathbf{cau}} C$. From the assumption and the transitivity of $<_{CP1}$ it follows that $A <_{CP1} C$. Let now $A = \langle T_1, h_1\rangle$, $B = \langle T_2, h_2\rangle$, and $C = \langle T_3, h_3\rangle$ and let us prove that $\langle T_3, h_1\rangle$ is an argument structure but $\langle T_1, h_3\rangle$ is not. Given $A \succ_{\mathbf{cau}} B$ we have that $\langle T_2, h_1\rangle$ is an argument structure (but not $\langle T_1, h_2\rangle$), and given $B \succ_{\mathbf{cau}} C$ we have that $\langle T_3, h_2\rangle$ is an argument structure too (but not $\langle T_2, h_3\rangle$). Note now that, by definition of argument structure, T_2 is minimal for the defeasible derivation of h_1, as well as T_3 is minimal for the defeasible derivation of h_2. That implies that $\langle T_2, h_1\rangle$ (resp. $\langle T_3, h_2\rangle$) cannot be a proper subargument of $\langle T_1, h_1\rangle$ (resp. $\langle T_2, h_2\rangle$), and since $\langle T_1, h_2\rangle$ (resp. $\langle T_2, h_3\rangle$) is not an argument structure, it follows that $\Pi \cup \{h_2\} \vdash h_1$ (resp. $\Pi \cup \{h_3\} \vdash h_2$). Therefore, $\Pi \cup \{h_3\} \vdash h_1$, which by Lemma 2 implies that (ii) $\langle T_3, h_1\rangle$ is an argument structure. On the other hand, by way of contradiction assume $\langle T_1, h_3\rangle$ is an argument structure. Since $\Pi \cup \{h_3\} \vdash h_2$, then we have that $\langle T_1, h_2\rangle$ is an argument structure. But this contradicts that $\langle T_1, h_1\rangle \succ_{\mathbf{cau}} \langle T_2, h_2\rangle$. Then $\langle T_1, h_3\rangle$ is not an argument structure. □

If an argument A is strictly more cautious than an argument B, then A will be considered a defeater of B (as well as of all the arguments of which B is a subargument), since A constitutes a proof of B not being maximally specific.

Definition 12 *(Cautious defeater).* Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. and let $\langle T_1, h_1\rangle$ and $\langle T_2, h_2\rangle$ be two argument structures obtained from $\mathcal{P}$. $\langle T_1, h_1\rangle$ is a *cautious defeater* of $\langle T_2, h_2\rangle$ iff there exists some subargument $\langle S, j\rangle$ of $\langle T_2, h_2\rangle$ such that $\langle T_1, h_1\rangle \succ_{\mathbf{cau}} \langle S, j\rangle$. Given a set of argument structures S we also define $def_{cau}(S) =_{df} \{(A, B) : A, B \in S$ and A is a cautious defeater of $B\}$.

Example 12 *(Example 4 revisited).* Add to the context of Example 4 the information that MS employees who are millionaires tend to be members of the MS Millionaires Society, expressed through the defeasible rule

$$ms\_millionaires\_society(x) \multimap ms\_employee(x), millionaire(x).$$

Then we have the argument

$$\begin{aligned} A' = \langle \{&millionaire(Beth) \multimap ms\_employee(Beth),\\ &ms\_millionaires\_society(x) \multimap ms\_employee(x), millionaire(x)\},\\ &ms\_millionaires\_society(Beth)\rangle, \end{aligned}$$

of which A is a subargument. Then C is an cautious defeater of both A and A' (Fig. 4).

Example 13 *(Example 6 revisited).* C is not at least as cautious as A, hence it is not a cautious defeater of A (note that both the defeasible rules of C and A enable the derivation of $flies(Tweety)$). (Nevertheless, A is not maximally specific since B is a proper defeater of it.)

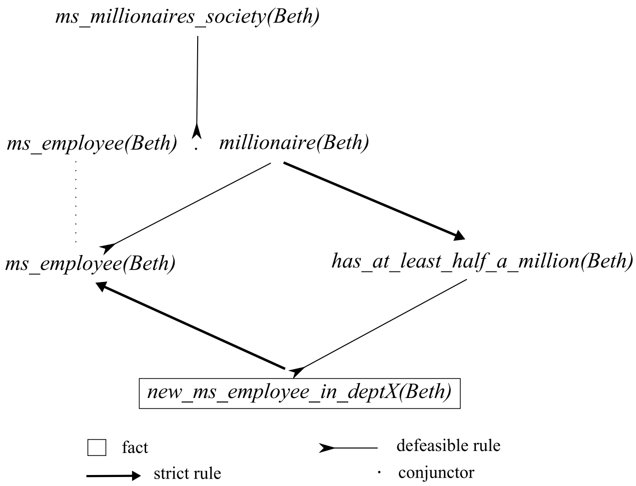

Fig. 4. De.l.p. of Example 12.

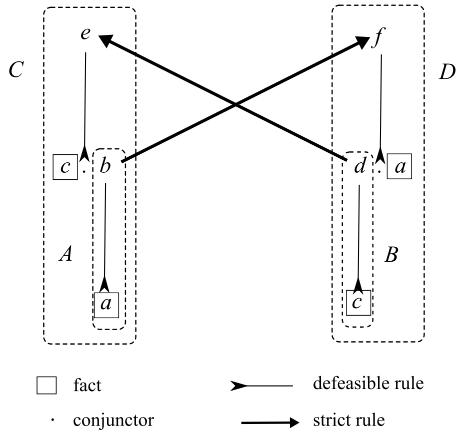

Fig. 5. De.l.p. of Example 15.

Example 14. Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. where $\Pi = \{a, b \leftarrow a, c \leftarrow d\}$ and $\Delta = \{d \prec b, e \prec d, c \prec a\}$. Then we have the arguments:

$$A = \langle\{c \prec a\}, c\rangle,$$

$$B = \langle\{d \prec b, e \prec d\}, e\rangle.$$

Then A is a cautious defeater of B since A is strictly more cautious than the subargument $\langle\{d \prec b\}, d\rangle$ of B, but A is not strictly more cautious than B (note that $\langle\{d \prec b, e \prec d\}, c\rangle$ is not an argument structure since it does not satisfy the minimality condition).

Example 15. Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. where $\Pi = \{a, c, f \leftarrow b, e \leftarrow d\}$ and $\Delta = \{b \prec a, d \prec c, e \prec b, c, f \prec a, d\}$. Then we have the arguments:

$$A = \langle\{b \prec a\}, b\rangle,$$

$$B = \langle\{d \prec\ < c\}, d\rangle.$$

$$C = \langle\{b \prec a, e \prec b, c\}, e\rangle,$$

$$D = \langle\{d \prec c, f \prec a, d\}, f\rangle.$$

Then C is a cautious defeater of D and vice versa. To see that C is a cautious defeater of D note that B is a subargument of D and $C \succ_{\mathbf{cau}} B$. On the other hand, D is a cautious defeater of C because A is a subargument of C and $D \succ_{\mathbf{cau}} A$ (Fig. 5).

Theorem 3. *The relation def_{cau} is irreflexive, non-symmetric, non-asymmetric, and non-transitive.*

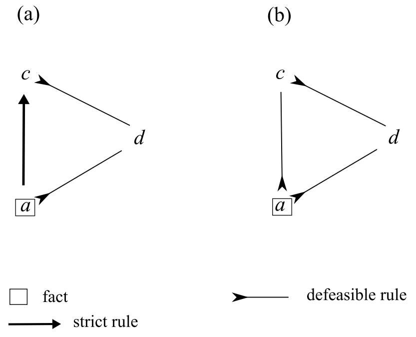

Fig. 6. Contrasting cases regarding cautiousness and specificity.

Proof. That def_{cau} is irreflexive and non-symmetric follows immediately from the fact that $\succ_{\mathbf{cau}}$ is irreflexive and non-symmetric. Example 15 shows that it is non-asymmetric and non-transitive. □

Being non-symmetric, non-asymmetric, and non-transitive, the relation def_{cau} can have cycles of any length, with the exception of cycles of length 1 (i.e. self-defeating arguments) due to the irreflexive property.
Naturally, we have that if argument A is at least as cautious as argument B, then A is at least as specific as B.

Proposition 1. $\succeq_{\mathbf{cau}} \subseteq \preceq_{CP1}$.

Proof. Immediate from definition.

To see that the converse is not true consider the following example.

Example 16. Let $\mathcal{P} = (\Pi, \Delta)$ where $\Delta = \{d \prec a, d \prec c\}$ and $\Pi = \{a, c \leftarrow a\}$ (Fig. 6 (a)). Then we have the arguments $A = \langle \{d \prec a\}, d\rangle$ and $B = \langle \{d \prec c\}, d\rangle$, which are such that $A <_{CP1} B$ but $A \not\succeq_{\mathbf{cau}} B$.

More importantly, exchanging above the strict rule with a defeasible rule we get an interesting contrast.

Example 17. Let $\mathcal{P} = (\Pi, \Delta)$ where $\Delta = \{c \prec a, d \prec a, d \prec c\}$ and $\Pi = \{a\}$ (Fig. 6 (b)). Then we have the arguments $A = \langle \{d \prec a\}, d\rangle$ and $B = \langle \{c \prec a, d \prec c\}, d\rangle$, which are such that $A \approx_{CP1} B$ and $A \approx_{\mathbf{cau}} B$ (cf. Example 3 and Example 9 in [37]. Moreover, note that $A \succ_{\mathbf{spec}} B$. More on this in Section 5.3).

3.4. Procedure for finding cautious defeaters

As we have argued throughout this paper, only maximally specific arguments should be justified, which – according to Definition 8 – implies that the mere existence of either a proper or a cautious defeater is sufficient for deeming the argument as not justified. Hence, the procedure for the justification of maximally specific arguments can take in part as a sub-procedure a procedure for finding proper defeaters (if a proper defeater is found then the argument is discarded). Moreover, another sub-procedure for finding cautious defeaters is needed.

We now outline a procedure for finding some cautious defeater for an input argument $\langle T, h\rangle$. We want the procedure to give an answer YES or NO and, in case of an affirmative answer, to return the item found. The argument can be represented by a *defeasible inference tree* in the style of [33], which consists of nodes and edges. The root node is h. The edges represent rules, so that each non-root node stands for a literal that belongs to the body of a rule, being the parent node the head of the rule. The leaves of the tree can only be matched with facts. More formally:

Definition 13. The *defeasible inference tree* of an argument structure $\langle T, h\rangle$, $\mathtt{T}_{\langle T,h\rangle}$, is a rooted tree where every node is labeled with a literal, satisfying the following conditions:

1. the root is labeled with h,
2. for each non-leaf node labeled with a literal L, there is a defeasible rule $L \prec L_1, \ldots, L_k \in T$ or a strict rule $L \leftarrow L_1, \ldots, L_k \in \Pi_G$, such that $L_1, \ldots, L_k$ are the labels of its children,
3. all the leaves are labeled with facts in Π_F.

Clearly, every proper subargument $\langle T', h'\rangle$ of $\langle T, h\rangle$ has its own defeasible inference tree, $\mathtt{T}_{\langle T',h'\rangle}$, which is a subtree of $\mathtt{T}_{\langle T,h\rangle}$. Let us call it a 'subargument tree'. Then, for each proper subargument $\langle T', h'\rangle$ we have to check whether there exists some argument which is strictly more cautious. To do that, a forward chaining through strict rules is run starting at the root node h' of the first subargument tree. If a rule $h'' \leftarrow h'$ is found but the rule $h' \leftarrow h''$ is not (according to Lemma 2, this checks in part conditions 2 and 3 of Definition 9), then a defeasible argument $\langle T'', h''\rangle$ is searched for such that $\langle T'', h''\rangle >_{CP1} \langle T', h'\rangle$. If such an argument does not exist or it is not strictly more cautious than $\langle T', h'\rangle$, then a new rule $h''' \leftarrow h''$ is searched for, and so on until either no further rule is found or a strictly more cautious argument is found. While no more cautious arguments are found the process continues with the next subarguments.

With regards to checking $\langle T'', h''\rangle >_{CP1} \langle T', h'\rangle$, the process is exponential in the number of literals of $\mathcal{T}_\Pi$, since the possible activation sets of an argument are all the elements of $Pow(\mathcal{T}_\Pi)$. A more efficient way of checking specificity is given by $CP2$, a slight modification of $CP1$ proposed by Wirth and Stolzenburg [37]. But though $CP1$ and $CP2$ coincide in most relevant cases, they are strictly not comparable. Since reducing the complexity of computing $CP1$ is beyond the scope of the present work, we just focus on the process of checking conditions 2 and 3 of $\succ_{\mathbf{cau}}$ (Definition 9).

To describe more details of the process consider again Example 12. Let

$$ms\_millionaires\_society(Beth)?$$

be the query to the program, for which the candidate argument

$$\begin{aligned}A' = \langle\{&millionaire(Beth) \prec ms\_employee(Beth),\\ &ms\_millionaires\_society(Beth) \prec ms\_employee(Beth), millionaire(Beth)\},\\ &ms\_millionaires\_society(Beth)\rangle\end{aligned}$$

is found. Then the process for finding a cautious defeater starts looking for a strict rule with body

$$ms\_millionaires\_society(Beth).$$

Since there exists no such a rule, no strictly more cautious argument for the (trivial) subargument A' is found. Then a new subargument is going to be identified, backtracking on the defeasible inference tree of A'. The rule

$$ms\_millionaires\_society(Beth) \prec ms\_employee(Beth), millionaire(Beth)$$

leads to find the proper subargument:

$$A'' = \langle\{millionaire(Beth) \prec ms\_employee(Beth)\}, millionaire(Beth)\rangle.$$

Then a strict rule with body *millionaire*$(Beth)$ is searched by forward chaining, finding the rule

$$has\_at\_least\_half\_a\_million(Beth) \leftarrow millionaire(Beth),$$

while a forward chaining on strict rules from $has\_at\_least\_half\_a\_million(Beth)$ does not lead to find $millionaire(Beth)$ (condition 2 and part of condition 3 of Definition 9 are checked). Next, a defeasible argument supporting

$has\_at\_least\_half\_a\_million(Beth)$ is constructed:

$$\begin{aligned}B = \langle &has\_at\_least\_half\_a\_million(Beth) \prec\\ &\quad new\_ms\_employee\_in\_deptX(Beth)\},\\ &\quad\quad has\_at\_least\_half\_a\_million(Beth)\rangle\end{aligned}$$

(this completes the checking of condition 3). Finally, the subset relationship between the activators of B and A'' is established (checking condition 1), verifying that the only activator $\{new\_ms\_employee\_in\_deptX(Beth)\}$ of B is also an activator of A'', while $\{ms\_emplpoyee(Beth)\}$ is an activator of A'' but not of B. Hence, a cautious defeater of A'', B, is found.

The procedure is sound, in the sense that if it gives a positive answer then it has found a cautious defeater.

Proposition 2. *If the procedure gives an affirmative answer for a given argument, then that argument has a cautious defeater.*

Proof. Assume the procedure gives an affirmative answer for $\langle T, h\rangle$. Then a subargument $\langle T', h'\rangle$ of $\langle T, h\rangle$ is found (following the defeasible derivation tree) such that (i) there exists a sequence of strict rules $h_n \leftarrow \ldots \leftarrow h_1$, $h_1 = h'$, and (ii) there exists no strict rule $h_{n-1} \leftarrow h_n$. By (i) and Lemma 2, we have that (1) $\langle T', h_n\rangle$ is an argument structure. Moreover, the procedure has also found (2) an argument $\langle T'', h_n\rangle$ such that $\langle T'', h_n\rangle <_{CP1} \langle T', h'\rangle$. Finally, by (ii) and the minimality condition on argument structures (i.e. condition 3 in Definition 1) we have that (3) $\langle T'', h'\rangle$ is not an argument structure. Then, from (1), (2), (3) and Definition 9 it follows that $\langle T'', h_n\rangle$ is more cautious than $\langle T', h'\rangle$. Therefore, by Definition 12, $\langle T'', h_n\rangle$ is a cautious defeater of $\langle T, h\rangle$.

On the other hand, note that the procedure is not able to find *every* cautious defeater in general. Despite this, if we want to check whether a particular argument $\langle S, j\rangle$ is or not a cautious defeater of another argument $\langle T, h\rangle$, we can think about modifying the procedure as to find a sequence of strict rules $h_n \leftarrow \ldots \leftarrow h_1$, where $h_1 = h$ and $h_n = j$, such that $h_{i-1} \leftarrow h_i$ is not found for some i, $1 < i \leq n$. If the result is negative, then $\langle S, j\rangle$ is not a cautious defeater of $\langle T, h\rangle$, since either $\langle T, j\rangle$ is not an argument structure or $\langle S, h\rangle$ is an argument structure (cf. Definition 9, conditions 2 and 3). Otherwise, if the result is positive the procedure goes ahead checking if $\langle S, j\rangle <_{CP1} \langle T, h\rangle$ (Definition 9, condition 1). Finally, if this is verified then the answer is affirmative, otherwise it is negative.

As said before, a process for checking $<_{CP1}$ is assumed. The searching for any cautious defeater also requires, as we have seen, a subprocess to find a sequence of chained strict rules such that the last one is not involved in a loop. The search space is a directed graph where nodes represent (possibly conjunctions of) literals and the directed arcs represent strict rules. To see an example, assume we want to check if $\langle T, h\rangle$ has a cautious defeater, and assume that the strict rules of the program are $j \leftarrow h$, $h \leftarrow j$, and $k \leftarrow j$. Then, after matching h with the body of the rule $j \leftarrow h$, the procedure will get the sequence of nodes $h - j - k$ in the best case, and the sequence $h - j - h - j - k$ in the worst case. But the search can be made more efficiently with the help of the graph-theoretical notion of *strongly connected components* (SCCs) of a graph, which are the equivalence classes of nodes under the relation of path-equivalence. In the above example, the SSCs of the graph $\langle\{h, j, k\}, \leftarrow\rangle$ are $\{h, j\}$ and $\{k\}$. These, in time, together with $\leftarrow$ form another graph with the only arc $\{k\} \leftarrow \{h, j\}$. Then, after identifying $\{h, j\}$ as the equivalence class of h we only have to find in this new graph a SCC S such that $S \leftarrow \{h, j\}$, which will give us $S = \{k\}$. In sum, this search will find the only sequence $\{h, j\} - \{k\}$. Note, on the other hand, that the search of the SCCs to get this new graph should be needed only once, and there exist algorithms that do that in linear time (see, e.g., [10]).

4. Filtering out non-maximally specific arguments through a dialectical analysis of warrant

Since non-maximally specific arguments are not acceptable, our proposal is to filter them out so that only maximally specific arguments can compete for warrant. In that competence reinstatement is valid, since the arguments interact through other kinds of defeaters than proper or cautious defeaters (for instance, blocking defeaters, as we will see below).

In DeLP, warrant can be determined through a dialectical analysis that can be represented by a two-party game, where a proponent tries to defend an argument and an opponent tries to refute it. We define a similar game in the terms used in [6]. Nevertheless, the arguments warranted in DeLP do not correspond in general to any of Dung's standard semantics [8,9]. But as Dung's semantics are based on acceptability, they are paradigmatic examples of semantics satisfying reinstatement (cf. [1]). This leads us to deviate from the warrant process of DeLP to use the formalism of Dung's argumentation frameworks. To do this, we take the set *Args* of all the argument structures that can be constructed in a de.l.p. $\mathcal{P}$ and all the defeat relations over *Args* are established to define the attack relation.

Two kinds of defeaters are defined in the simplest version of DeLP: proper defeaters and blocking defeaters. In our approach we add cautious defeaters with the aim of rejecting non-maximally specific arguments. Moreover, other defeat criteria could be added obtaining more complex systems. Since, as we have argued, all the criticisms to reinstatement are saved (we hope) with our redefinition of the specificity criterion, no reason remains in principle to think that new defeat criteria might invalidate reinstatement. If this is accepted, then reinstatement would be sound in a system which combines different kinds of defeat criteria whenever proper and cautious defeaters are also considered for the plain rejection of non-maximally specific arguments. The only caveat is that 'proper defeater' is considered to be defined in terms of $<_{CP1}$ instead of $\succ_{\mathbf{spec}}$, as we have seen in section 3.

Definition 14 *(Argumentation framework associated to a de.l.p.).* Given a de.l.p. $\mathcal{P}$, the *argumentation framework associated to* $\mathcal{P}$ is the pair $(Args, attacks)$ where *Args* is the set of all the argument structures obtained from $\mathcal{P}$ and $attacks = \bigcup DEF(Args)$, where $DEF(Args) = \{def_1, \ldots, def_k\}$, is the set containing all the defeat criteria $def_i \subseteq Args \times Args$ $(1 \leq i \leq n)$ defined on *Args*.

We will assume that $def_{prop}(Args)$, $def_{cau}(Args) \in DEF(Args)$. Moreover, we also want to include blocking defeaters in $DEF(Args)$ as in the original presentation of DeLP. Again, we follow [18] but change $\succ_{\mathbf{spec}}$ to $<_{CP1}$.

Definition 15 *(Blocking defeater).* Given a de.l.p. $\mathcal{P} = (\Pi, \Delta)$, $\langle T, h\rangle$ is a *blocking defeater* of $\langle T', h'\rangle$ iff there exists some sub-argument $\langle T'', h''\rangle$ of $\langle T', h'\rangle$ such that $\Pi \cup \{h, h''\} \vdash \perp$, and $\langle T, h\rangle \not<_{CP1} \langle T'', h''\rangle$ and $\langle T'', h''\rangle \not<_{CP1} \langle T, h\rangle$. Given a set of argument structures S we also define $def_{block}(S) =_{df} \{(A, B) : A, B \in S$ and A is a blocking defeater of $B\}$.

Definition 16 *(Argumentation game).* An *argumentation game* on an argumentation framework $(Args, attacks)$ is a zero-sum extensive game in which:

1. There are two players, i and $-i$, who play the roles of **P** and **O**, respectively.
2. A *history* in the game is any sequence $A_0, A_1, A_2, \ldots, A_{2k}, A_{2k+1}, \ldots$ of choices of arguments in *Args* made by the players in the game. A_{2k} corresponds to **P** and A_{2k+1} to **O**, for $k = 0, 1, \ldots$. At any history, A_0 is the argument that player **P** intends to defend.
3. In a history, the choices by a player i at a level $k > 0$ are $C_i(k) = \{A \in Args : (A, A_{k-1}) \in attacks\}$.
4. A history of finite length K, $A_0, \ldots, A_K$, is *terminal* if A_K corresponds to player j ($j = i$ or $j = -i$) and $C_{-j}(K+1) = \emptyset$.

5. Payoffs are determined at terminal histories: at $A_0, \dots, A_K$, **P**'s payoff is 1 (representing winning) if K is even (*i.e.*, **O** cannot reply to **P**'s last argument), and -1 (representing loosing) otherwise. In turn, **O**'s payoff at $A_0, \dots, A_K$ is 1 if K is odd and -1 otherwise.

Definition 17 *(Strategy).* A *strategy* for a player i is a function that assigns an element $A_{l+1} \in C_i(l)$ at each non-terminal history $A_0, \dots, A_l$ where A_l corresponds to player $-i$. A strategy of player i is said a *winning strategy* for i if for every strategy chosen by $-i$, the ensuing terminal history yields a payoff 1 for player i.

Definition 18 *(Warrant).* An argument A is *warranted* in $(Args, attacks)$ iff **P** has a winning strategy to defend A in the game associated to $(Args, attacks)$.

So defined, the plain game enables the reinstatement of any argument (maximally specific or not) since no differential treatment is considered for attacks originated by different kinds of defeaters. On the other hand, special behaviors can be captured by defining specific protocols, i.e., additional rules stating further obligations, permissions, and victory conditions. In order to reject non-maximally specific arguments, the proposal is simple: once a player moves a proper or a cautious defeater the other party loses the game. Moreover, other features can be determined by the protocol. For instance, in order to capture a skeptical behavior the following rule can be stated:

- **P** is not allowed to repeat previous moves of either player

Intuitively, P is obliged not to repeat her moves in order to show that her proposal can ultimately be defended with a warranted, non-attacked argument. On the other hand, if P repeats O's moves then she is using an argument that attacks one of her own arguments. Let us call the protocol conformed to this only rule the 'skeptical protocol'. It is known that the skeptical protocol defines a game such that **P** has a winning strategy to defend A iff A belongs to the *grounded extension* of the framework [26]. Briefly, Dung [11] defines the *grounded extension* of an argumentation framework $(Args, attacks)$ as the least fixed point of the operator $F(\cdot)$ such that for every subset $S \subseteq Args$, $F(S) = \{A \in Args\colon \forall B \in Args((B, A) \in attacks \Rightarrow \exists C \in S(C, B) \in attacks)\}$ (i.e. the grounded extension is the least subset $S \subseteq Args$ such that $S = F(S)$).

In order to combine the exclusion of non-maximally specific arguments with a skeptical behavior we propose the following protocol:

Protocol 1. At any level k,

1. neither player is allowed to move if the argument played in level $k-1$ is either a proper or a cautious defeater of the argument played in level $k-2$;
2. **P** is not allowed to move an argument that was already advanced by either player in the same history.

The first rule says that once a proper or a cautious defeater is played the game ends (the player who moved the previous argument loses since she played a non-maximally specific argument). The purpose of this rule is, of course, to commit the players to use only maximally specific arguments. The second rule, in time, captures the skeptical behavior while ensures finite games.

The next is an interesting example also introduced by Horty [21] and discussed in the literature [30].

Example 18. Let $\mathcal{P} = (\Pi, \Delta)$ be the de.l.p. representing the following information: Larry is a public defender and is a tenant in Brentwood, all public defenders are lawyers, public defenders tend not to be rich, lawyers tend to be rich, Brentwood residents tend to be rich, and Brentwood tenants tend not to be rich.

$$\begin{aligned}
\Pi = \{ & lawyer(x) \leftarrow public_defender(x),\\
& brentwood_resident(x) \leftarrow brentwood_tenant(x),\\
& brentwood_tenant(Larry), public_defender(Larry)\}\\
\Delta = \{ & rich(x) \prec lawyer(x),\\
& \sim rich(x) \prec public_defender(x),\\
& rich(x) \prec brentwood_resident(x),\\
& \sim rich(x) \prec brentwood_tenant(x)\}
\end{aligned}$$

Then we have the argument structures:

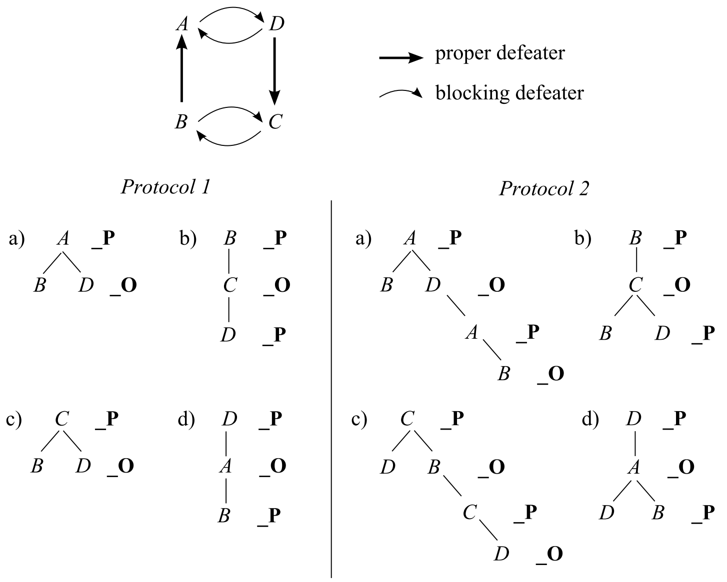

Fig. 7. Argumentation framework of Example 18 and tree-form representations of the games for defending each argument under Protocol 1 and Protocol 2.

$$A = \langle\{rich(Larry) \multimap lawyer(Larry)\}, rich(Larry)\rangle,$$

$$B = \langle\{\sim rich(Larry) \multimap public\_defender(Larry)\}, \sim rich(Larry)\rangle,$$

$$C = \langle\{rich(Larry) \multimap brentwood\_resident(Larry)\}, rich(Larry)\rangle,$$

$$D = \langle\{\sim rich(Larry) \multimap brentwood\_tenant(Larry)\}, \sim rich(Larry)\rangle.$$

We have $def_{prop} = \{(B, A), (D, C)\}$ and $def_{block} = \{(A, D), (D, A), (B, C), (C, B)\}$, with the resulting associated framework depicted in Fig. 7 (top). In the respective game, playing Protocol 1, **P** has winning strategies to defend both B and D, while **O** has winning strategies to refute both A and C (Fig. 7, left).

In accordance with the intuition, Protocol 1 enables to conclude that Larry is not rich. Arguments A and C are discarded. In the case of A, for instance, **O** has two possible winning strategies to respond: moving B or moving D. B is a proper defeater of A (this move shows that A is not maximally specific); and D could not be responded by **P** since the only argument attacking it is A, which was already played. In sum, the set of warranted arguments following Protocol 1 is $\{B, D\}$.

On the other hand, in the plain game (i.e. without any further protocol) no conclusion is obtained about whether Larry is rich or not, since neither argument results warranted (**P** cannot defend neither argument since the respective games would be infinite). The same results are obtained with the skeptical protocol. In the defense of A, for instance, **O** has two winning strategies yielding the histories $A - D$ and $A - B - C - B$, respectively (similarly, for D we have $D - A$ and $D - C - B - C$). These results coincide with the grounded extension $\emptyset$ of the argumentation framework.

The desired behavior regarding maximally specific arguments can also be achieved in combination with a credulous behavior. Vreeswijk and Prakken [36], for instance, have defined game rules in order to capture Dung's *preferred semantics* [11]. Briefly, given an argumentation framework $(Args, attacks)$, a *preferred extension* is a maximal (w.r.t. $\subseteq$) subset $S \subseteq Args$ such that (i) for every argument $A \in S$ and for every argument $B \in Args$ such that $(B, A) \in attacks$, there exists some argument $C \in S$ such that $(C, B) \in attacks$; and (ii) for every pair of arguments $A, B \in S$, $(A, B) \notin attacks$. Now, Vreeswijk and Prakken's protocol captures the defense of arguments belonging to preferred extensions with the following rules:

- neither player can repeat the arguments used by **O**,
- neither player is allowed to move if the last argument in the sequence was already played by **O**,
- if **O** is out of moves according to the above rules, then she can start a new line of attack on a previous argument of **P**.

Let us call the protocol conformed to these three rules the 'credulous protocol'. Then, adding to the credulous protocol the rule dictating the victory for the player who moves either a proper or a cautious defeater we obtain the following protocol:

Protocol 2. At any level k,

1. neither player is allowed to move if the argument played in level $k - 1$ is either a proper or a cautious defeater of the argument played in level $k - 2$;

2. neither player is allowed to move an argument used by **O** in a level $k' < k$;
3. neither player is allowed to move if the argument played in level $k-1$ is the same argument played by **P** in a level $k' < k$;
4. if **O** is out of moves according to the above rules, then she can start a new line of attack on an argument played by **P** in any level $k' < k$.

Let us see the games of Example 18 again, played this time under Protocol 2 (Fig. 7, right). Argument A (similarly, argument C) cannot be defended since **O** has two winning strategies yielding the sequences $A - B$ and $A - D - A - B$, respectively. On the other hand, argument B (similarly, argument D) can be defended by **P** this time with two winning strategies yielding the histories $B - C - D$ and $B - C - B$, respectively. Hence, in this example we obtain the same results as under Protocol 1, that is, the set of warranted arguments is $\{B, D\}$ (Larry is not rich).

Note that with the credulous protocol (i.e. with rules 2–4 only) all the arguments can be defended with winning strategies: $A - B - C - D - A$, $B - C - B$, $C - D - A - B - C$, and $D - A - D$, for A, B, C, and D, respectively, in accordance with the preferred extensions $\{A, C\}$ and $\{B, D\}$. In consequence, no conclusion can be obtained about Larry's richness.

We want to show now that, once all non-maximally specific arguments are removed from an argumentation framework, Protocol 1 is sound and complete with respect to grounded semantics, as is Protocol 2 with respect to preferred semantics. The argumentation framework resulting from removing all non-maximally specific arguments is defined as follows.

Definition 19. We define the *maximally specific part* of an argumentation framework $(Args, attacks)$ as the argumentation framework $MS(Args, attacks) = (Args_{MS}, attacks_{MS})$, where $Args_{MS} = Args \setminus \{A \in Args : \exists B \in Args((B, A) \in def_{prop} \vee (B, A) \in def_{cau})\}$, and $attacks_{MS} = \{(A, B) : A, B \in Args_{MS}, (A, B) \in attacks\}$.

The result we are looking for rests on the following two lemmata.

Lemma 3. *For every argumentation framework $(Args, attacks)$ and for every argument $A \in Args$,* **P** *has a winning strategy to defend A under the skeptical protocol iff A belongs to the grounded extension of $(Args, attacks)$.*

Proof. In [26] it is proved that A belongs to the grounded extension iff **P** has a winning strategy to defend A in which the set of arguments moved by **P** is conflict-free (i.e. $(A, B) \notin attacks$ for any pair of arguments A and B played by **P**) and no move repeats her own previous moves [26]: Theorem 6.3). Hence, we only have to prove that P's set of moves is conflict-free iff **P** does not repeat the moves by **O**. By contraposition, it is obvious that if **P** repeats an argument previously played by **O** then she introduces a conflict in her strategy, since every move by **O** is an argument attacking some argument chosen by **P**. On the other hand, if **P**'s strategy is not conflict-free then **O** can use a move of **P**, forcing **P** to follow her own steps. □

Lemma 4. *For every argumentation framework $(Args, attacks)$ and for every argument $A \in Args$,* **P** *has a winning strategy to defend A under the credulous protocol iff A belongs to some preferred extension of $(Args, attacks)$.*

Proof. This result is proved in [36] (Proposition 1). □

Proposition 3. *For every argumentation framework $(Args, attacks)$ and for every argument $A \in Args$,* **P** *has a winning strategy to defend A in the argumentation game on $(Args, attacks)$ under*

1. *Protocol 1 iff A belongs to the grounded extension of $MS(Args, attacks)$*
2. *Protocol 2 iff A belongs to some preferred extension of $MS(Args, attacks)$*

Proof. The proof is analogous for 1 and 2, hence we offer both proofs in a single version, divided in the "if" part and the "only if" part.

(if) Assume A belongs to the grounded extension (a preferred extension) of $MS(Args, attacks)$. By definition of $MS(Args, attacks)$, every argument used in the game on that argumentation framework is maximally specific. Now, by Lemma 3 (Lemma 4), **P** has a winning strategy to defend A under the skeptic (credulous) protocol in the game on $MS(Args, attacks)$. Then, since every argument used by **O** is maximally specific and **P** counts with a successful response for each of them, then **P** can also use the same strategy in the game on $(Args, attacks)$ to respond to all those attacks under Protocol 1 (Protocol 2). On the other hand, if **O** uses some non-maximally specific argument in the game on $(Args, attacks)$, then **P** counts with either a proper or a cautious defeater to finish successfully the game via the rule 1. Therefore, **P** has a winning strategy to defend A in the argumentation game on $(Args, attacks)$ under Protocol 1 (Protocol 2).

(only if) Assume **P** has a winning strategy to defend A in the argumentation game on $(Args, attacks)$ under Protocol 1 (Protocol 2). Then, by rule 1, every argument in that strategy is maximally specific, which means that all of them are free of both proper and cautious defeaters. Hence, this strategy can also be successfully used by **P** in the game on $MS(Args, attacks)$. Therefore, by Lemma 3 (Lemma 4), A belongs to the grounded extension (a preferred extension) of $MS(Args, attacks)$. □

This result shows that Protocols 1 and 2 are effective to remove non-maximally specific arguments without affecting the warrant process of other arguments.

5. Discussion

We have argued that the problems referred by Horty about reinstatement are indeed problems concerning how to define a specificity criterion. And we have also argued that our notion of maximal specificity can solve those problems. The point is that non-maximally specific arguments should be defeated outright and hence not able of being reinstated. Otherwise, as we have seen, the improper reinstatement of such arguments would interfere in the warrant process yielding unacceptable results as those imagined by Horty. We discuss this issue next. Furthermore, we will discuss possible relations among specificity and Besnard and Hunter's notion of 'conservativeness' [3,4] and comment related works on specificity.

5.1. Possible counterexamples?

Here we discuss some other examples that could suggest that some non-maximally specific arguments should be reinstated anyway. For instance, consider again Example 2 but replacing argument C with C': "It cannot be concluded that Tweety is a penguin since it was observed under deficient sight conditions during a blizzard" (example due to Prakken [30]). Then C' could be seen as an undercutting defeater of B and B as a proper defeater of A, what would lead to the reinstatement of the non-maximally specific argument A. But note that the acceptance of C' implies the treatment of 'Tweety is a penguin' not as evidence but as a questionable assumption. Hence, B should not be treated as a proper defeater of A strictly. Therefore, A is still a maximally specific argument and its reinstatement seems right. In agreement with Prakken, we have to admit that there exists a problem of representation here. Proper defeat is defined for arguments built up from evidence, not from assumptions.

On the other hand, in agreement with Prakken again, we have to admit that there exists a representational issue which is evident in some cases. Let us replace C' above with C'': "Tweety flies since he is equipped with a personal flying device". We will have different results depending on how this information is represented. If the defeasible rule is

$$flies(x) \multimap equipped\_with\_a\_flying\_device(x)$$

then we have a blocking defeater of B, but if the rule is instead

$$flies(x) \multimap penguin(x), equipped\_with\_a\_flying\_device(x)$$

then we have a proper defeater of B. In the first case, our formalism will skeptically (Protocol 1) reject all the three arguments, and credulously (Protocol 2) accept alternatively arguments B and C''. In the second case, instead, we have both a skeptical and a credulous acceptance of argument C'' only. Nevertheless, note that argument A cannot be reinstated in either case, while in other formalisms that do not outrightly reject non-maximally specific arguments A can be reinstated in both cases. This means that our approach is capable of mitigating, at least to some extent, the representational issue.

Prakken [30] makes a distinction between two kinds of reinstatement: direct and indirect. Direct reinstatement is when all the arguments are in conflict on their final conclusions (e.g. Example 2). Indirect reinstatement, on the other hand, is when the reinstating argument defeats the 'middle' argument on one of its intermediary conclusions (e.g. Example 3). Unlike direct reinstatement – Prakken argues – indirect reinstatement is valid. This distinction is not related to our solution as it does not focus on the kinds of defeaters which are involved and the role of undermining evidence on them, which is the key for determining maximal specificity. In our view, indeed, indirect reinstatement is invalid for cautiously defeated arguments.

The following example was also introduced by Prakken to argue that reinstatement depends "on the nature of the domain, the kind of knowledge involved and the context in which this knowledge is used" [30]: 93):

Example 19.

- A: John will be imprisoned up to 6 years because for theft imprisonment up to 6 years is acceptable, and John has been found guilty of theft.
- B: John will be imprisoned for no more than 3 years because for theft out of poverty imprisonment of more than 3 years is not acceptable, and evidence shows that John stole motivated by poverty.
- C: John will be imprisoned for more than 4 years because he stole during riots, and for theft during riots, even when poverty is proved, only imprisonment of more than 4 years is acceptable.

Prakken argues that the reinstatement of A by C is valid here and leads to accept an imprisonment between 4 and 6 years. We disagree at this point since, anyway, C seems to be a proper defeater of B and B a proper defeater of A, hence both A and B should be rejected as they are non-maximally specific. The total evidence considered in C about a more serious crime than theft out of poverty leads to put a minimum of 4 years of imprisonment, leaving the upper limit not established. Indeed, we can imagine even more serious crimes (e.g. murder) which occurrence together with theft would rise the top above 6 years. Hence, we think that C should be the only warranted argument.

Though our notion of maximal specificity is not completely characterized, since we stated some necessary but not sufficient conditions, the approach here improves that of [5]. We have identified cautious defeaters which make appear some arguments based on partial evidence as fallacious. Example 4 is clear about that, where the argument for Beth having at least half a million is not an undermining defeater but a cautious defeater of the argument for Beth being a millionaire. Another one is Example 5, though in that case the issue of representation arises again: the situation cannot be represented in DeLP because disjunctions cannot occur in the head of a rule. Anyway, the information can be adjusted to the DeLP formalism by considering that Green Island and Sand Island form a group of islands, call it 'Two Islands'. We obtain the following representation:

Example 20. Let $\mathcal{P} = (\Pi, \Delta)$ be a de.l.p. representing the knowledge that all least ruffed finches are ruffed finches, ruffed finches tend to nest on Green Island, least ruffed finches tend to nest on Two Islands, nesting on Green Island implies nesting in Two Islands, and Frank is a least ruffed finch:

$$\begin{aligned}\Pi = \{&\mathit{ruffed\_finch}(x) \leftarrow \mathit{least\_ruffed\_finch}(x),\\ &\mathit{nests\_on\_TwoIslands}(x) \leftarrow \mathit{nests\_on\_GreenIsland}(x),\\ &\mathit{least\_ruffed\_finch}(\mathit{Frank})\}\\ \Delta = \{&\mathit{nests\_on\_GreenIsland}(x) \prec \mathit{ruffed\_finch}(x),\\ &\mathit{nests\_on\_TwoIslands}(x) \prec \mathit{least\_ruffed\_finch}(x)\}\end{aligned}$$

Then we have, among others, the argument structures:

$$\begin{aligned}A = \langle\{&\mathit{nests\_on\_GreenIsland}(\mathit{Frank}) \prec \mathit{ruffed\_finch}(\mathit{Frank})\},\\ &\qquad \mathit{nests\_on\_GreenIsland}(\mathit{Frank})\rangle\\ B = \langle\{&\mathit{nests\_on\_TwoIslands}(\mathit{Frank}) \prec \mathit{least\_ruffed\_finch}(\mathit{Frank})\},\\ &\qquad \mathit{nests\_on\_TwoIslands}(\mathit{Frank})\rangle\end{aligned}$$

Though $B \succ_{\mathbf{spec}} A$ (and also $B <_{CP1} A$), B is not a proper defeater of A, hence A – and the conclusion that Frank nests on Green Island – is warranted in the original formulation of DeLP. The formalism incurs in the fallacy of exclusion, since the information that Frank is a least ruffed finch is obviated, treating Frank just as a ruffed finch. But A is not maximally specific and should be rejected, as it is the case considering cautious defeaters. Horty agrees about that rejection, but proposes to add a (meta-level) default expressing that cases of least ruffed finches exclude the application of the default that connects ruffed finches with nesting on Green Island (a kind of undercutting defeater).

5.2. Is maximal specificity related to conservativeness?

An interesting question is how the notion of maximal specificity is related to the seemingly close Besnard and Hunter's notion of *conservativeness* [3,4]. The intuition behind conservativeness is that, "[r]oughly speaking, a more conservative argument is more general: It is, so to speak, less demanding on the support and less specific about the consequent." [4]: 42). In other words, we can say that a more conservative argument uses a more general support to derive a more general conclusion. A maximally specific argument, on the other hand, is more general (more "conservative") with respect to the conclusion because of a more specific support. That is, more specific reasons lead to be more *cautious* about the conclusion. The difference is clear for at least two different meanings of 'support'. If it refers to the possible information which makes applicable the rules of the arguments (i.e. the activators) then, in the above Example 20, A is not more conservative than B, since being a ruffed finch is more general than being a least ruffed finch but nesting in Green Island is not more general than nesting in Two Islands (it is more specific, indeed). On the other hand, if 'support' is intended to mean all the information considered in an argument, that is, both the activators and the rules, these notions cannot be subsumed one by the other neither. Indeed, this last is the meaning we understand from Besnard and Hunter's formulation. Let us make a more formal analysis.

Conservativeness is introduced as a relation among deductive arguments in a setting formalized in terms of classical logic. A fixed, possibly inconsistent set Δ of formulae is assumed as representing a repository of information from which arguments can be constructed. In this setting, an *argument* is a pair $\langle\Phi, \alpha\rangle$ such that 1) $\Phi \nvdash \perp$, 2) $\Phi \vdash \alpha$, and 3) Φ is a minimal subset of Δ satisfying 2. Then, an argument $\langle\Phi, \alpha\rangle$ is *more conservative* than an argument $\langle\Psi, \beta\rangle$ iff $\Phi \subseteq \Psi$ and $\{\beta\} \vdash \alpha$. For example, $\langle\{p, p \to q\}, q\rangle$ is more conservative than $\langle\{p, p \to q, q \to r\}, r \wedge q\rangle$ [4]: 42). Beyond the obvious fact that our setting is not applied to deductive arguments, we will try a comparison with our notion of maximal specificity through the last Example 20. Let us try a representation in Besnard and Hunter's setting. Remember that all formulae are treated equitably and they all belong to the "repository" Δ.

Example 21. Let

$$\begin{aligned}\Delta = \{\, & least\_ruffed\_finch(Frank) \rightarrow ruffed\_finch(Frank),\\ & nests\_on\_GreenIsland(Frank) \rightarrow nests\_on\_TwoIslands(Frank),\\ & ruffed\_finch(Frank) \rightarrow nests\_on\_GreenIsland(Frank),\\ & least\_ruffed\_finch(Frank) \rightarrow nests\_on\_TwoIslands(Frank)\}\\ & least\_ruffed\_finch(Frank)\}\end{aligned}$$

Then we have the arguments

$$\begin{aligned}A = \langle\{\, & least\_ruffed\_finch(Frank),\\ & least\_ruffed\_finch(Frank) \rightarrow ruffed\_finch(Frank),\\ & ruffed\_finch(Frank) \rightarrow nests\_on\_GreenIsland(Frank)\,\},\\ & nests\_on\_GreenIsland(Frank)\ \rangle\end{aligned}$$

and

$$\begin{aligned}C = \langle\{\, & least\_ruffed\_finch(Frank),\\ & least\_ruffed\_finch(Frank) \rightarrow nests\_on\_TwoIslands(Frank)\,\},\\ & nests\_on\_TwoIslands(Frank)\ \rangle.\end{aligned}$$

which are not related by conservativeness. Note in particular that the corresponding supports are not related by $\subseteq$.

5.3. Related approaches to specificity

As we have claimed before, our notion of maximal specificity is partially characterized since we have stated some necessary but not sufficient conditions. This means that further conditions can be found. At this point, we have to consider other works which have discussed the original version of specificity by Poole [29] in the context of argumentation systems. Basically, they are the works by Stolzenburg, García, Chesñevar, and Simari [34] and Wirth and Stolzenburg [37]. Moreover, we want to compare our intuitions with some insights of Dung and Son [12]. On the other hand, works on specificity related to inheritance networks, default logics and other non-monotonic reasoning systems – which have been widely discussed in the literature for at least thirty-five years – are not included this discussion.

5.3.1. Stolzenburg et al. [34]

The work by Stolzenburg et al. [34] deals with a problem in the formalization of the specificity relation, as originally defined by Poole and used in DeLP. The problem is that the comparison of two arguments by specificity is made in terms of the activation sets of both arguments, and some rules -strict or defeasible- which are not used in the arguments at stake can provoke undesired activations; thus, the comparison may not yield the desired outcome. Consider the following example ([34]: Example 11):

Example 22. Let $\mathcal{P} = (\Pi, \Delta)$ where $\Delta = \{(x \prec a, b, c), (\sim x \prec a, b), a \prec d, b \prec e, f \prec e\}$ and $\Pi = \{c, d, e, (x \leftarrow a, f)\}$. Then we have the arguments $A = \langle\{(x \prec a, b, c), a \prec d, b \prec e\}, x\rangle$, $B = \langle\{(\sim x \prec a, b), a \prec d, b \prec e\}, \sim x\rangle$.

Intuitively, A should be more specific than B, but this is not the case since $\{d, f\}$ is an activation set of A (because of the strict rule $(x \leftarrow a, f)$) but it is not an activation set of B. The solution is achieved by characterizing the class of *non-trivial activation sets* of an argument. To do this, the *completions* of the arguments are considered. Roughly, a completion is a set of rules – strict or defeasible – that are used in the argument. Then, the non-trivial activation sets of an argument are subsets of literals occurring in the head or the body of each rule in the completion, that non-trivially derive the conclusion of the argument. Hence, $\{d, f\}$ is not a non-trivial activation set for A, because f does not occur in the derivation which uses only defeasible rules of A. The flaws detected by Stolzenburg et al. with respect to Poole's specificity are inherited by DeLP's specificity, and are overcome by Wirth and Stolzenburg's $CP1$ [37].

5.3.2. Wirth and Stolzenburg [37]

$CP1$ is the transitive version of specificity that we have used to characterize our notions of maximal specificity and cautious defeaters. The problem of not satisfying transitivity is crucial if specificity is understood as a preference relation used as a comparison criterion for argument choice. Under this perspective, if we have three arguments, A, B and C, such that A is more specific than B and B is more specific than C, we need A being more specific than C, otherwise the choice of A could be disrupted by the interference of C. We have followed Hempel's intuition of taking the maximally specific arguments, discarding the rest. But maximal specificity is not decided just taking the maximal elements of a single relation, but different relations on arguments (like those behind proper defeaters, cautious defeaters, or even others[9]) can lead to deem an argument as not maximally specific.

$CP1$ also satisfies *monotonicity w.r.t. conjunction*, roughly, the property by which if two arguments A and B are respectively more specific than C and D, then the conjunction of A and B should be more specific than the conjunction of C and D. This property is intuitive but it is not satisfied by Poole's specificity. Consider the following example ([29]: Example 6, and [37]: Example 12):

Example 23. Let $\mathcal{P} = (\Pi, \Delta)$ where $\Delta = \{b \prec a, c \prec b, e \prec d, f \prec e, \sim c \prec a, \sim f \prec d\}$ and $\Pi = \{a, d, (g_1 \leftarrow \sim c, \sim f), (g_2 \leftarrow c, f)\}$. Then we have the arguments $A_1 = \langle \{\sim c \prec a, \sim f \prec d\}, g_1\rangle$ and $A_2 = \langle \{b \prec a, c \prec b, e \prec d, f \prec e\}, g_2\rangle$.

Poole introduced this example to show that, under his formulation, and contrarily to his intuition, A_1 and A_2 are not comparable by specificity, though $\langle \{\sim c \prec a, \sim c\rangle$ and $\langle \{\sim f \prec d\}, \sim f\rangle$, which are subarguments of A_1, are respectively more specific than $\langle \{b \prec a, c \prec b, c\rangle$ and $\langle \{e \prec d, f \prec e\}, f\rangle$, which are subarguments of A_2. Nevertheless, in the setting of [37], A_1 and A_2 are equivalent with respect to $CP1$, as well as their respective subarguments. The difference lies in the *activation sets* taken into account to define each specificity relation: while Poole takes as activation sets the subsets of literals occurring in formulae of either Π or Δ, Wirth and Stolzenburg take subsets of literals occurring only in formulae of Π. In this way, the above example is not a counterexample for the monotonicity w.r.t. conjunction of $CP1$. Consider now the behavior of our criterion of maximal specificity. $\langle \{\sim c \prec a, \sim c\rangle$ and $\langle \{\sim f \prec d\}, \sim f\rangle$ are, respectively, proper defeaters of $\langle \{b \prec a, c \prec b, c\rangle$ and $\langle \{e \prec d, f \prec e\}, f\rangle$. Then, A_2 is not maximally specific. On the other hand, A_1 is maximally specific. In sum, in this example we get the outcome desired (but not achieved) by Poole, though it is at odds with Wirth and Stolzenburg's intuition.

Another source of doubt about Wirth and Stolzenburg's $CP1$ relation, in our view, is the treatment of the argumentation framework depicted in Fig. 6 (b). For the authors, $\langle \{d \prec a\}, d\rangle$ is not "more concise" than $\langle \{d \prec c, c \prec a\}, d\rangle$, but it is so when $c \leftarrow a \in \Pi$ (cf. also Examples 3 and 9 in [37]). However, this is at odds with the opinion that seems to be more widely shared, and that can be tracked back in the history of defeasible reasoning, expressed in terms as *shortest-path* [16], *directness* [24], and even in the specificity ideas of Poole [29], Simari and Loui [33], and García and Simari [18]. In sum, while $CP1$ solves the problem of transitivity, it still yields some behaviors that seem to contrast with common sense.

5.3.3. Dung and Son [12]

Another approach to reasoning with specificity is that of Dung and Son [12]. These authors define a *more specific* relation among defaults which combines requirements on the respective activators, similarly to the definition of $\succ_{\mathbf{spec}}$, with a requirement of conflict between the defaults. Intuitively, this comparison criterion is closer to the notion of proper defeater than to $\succ_{\mathbf{spec}}$. This is because the existence of conflict is considered necessary for the comparison. Moreover, these authors seem to think that such a relation of "specificity-with-conflict" is also sufficient for a comparison by specificity, as they say that "a more specific relation among non-conflicting pairs of defaults would be spurious" [12]: 43). In consequence, the problems approached through cautious defeaters in our work are disregarded in the approach by Dung and Son from their very conception of specificity.

This last comment also gives rise to the question of what kind of defeaters are cautious defeaters. We think they qualify as undercutting defeaters. As cautious defeaters are based on a total-evidence requirement, they can be considered a kind of – in Pollock's terms – subproperty defeaters, just the same as proper defeaters. And subproperty defeaters are all undercutting defeaters. Pollock's words seem to confirm our opinion:

> To the best of my knowledge, there has never been an intuitive example of specificity defeat presented anywhere in the literature that is not an example of the operation of the total-evidence requirement in one of these special varieties of defeasible inference [statistical syllogism, direct inference, various kinds of legal and deontic reasoning], and the latter are all instances of undercutting defeat. Accordingly, I will assume that undercutting defeaters and rebutting defeaters are the only possible kinds of defeaters [28]: 236).

5.4. Further comments about reinstatement and other principles

More in the line of Horty's solution to the reinstatement problem, the work by Modgil on hierarchical argumentation [25] offers another interesting turn to the problem, introducing arguments for (meta-level) preference criteria. The model develops a form of meta-argumentation where, for example, if A *attacks* B is established on basis of a preference criterion P1, and B *attacks* A is established on basis of a preference criterion P2, an argument C supporting the preference of P1 over P2 poses an attack on B *attacks* A, A resulting reinstated. Note that, under this view, C is not attacking B but *the attack* of B over A. The example of Tweety observed during a blizzard can be interpreted in these terms assuming that the preference

[9] Remember that Definition 8 states some necessary but not sufficient conditions.

criterion is based on an ordering $>$ on the evidence, such that $bird(Tweety) > penguin(Tweety)$. Then, while B is a proper defeater of A, C expresses a preference of A over B based on $>$, so that C defends A. Examples like Example 20, on the other hand, cannot be solved in this way unless, again, a special kind of undercutting defeater is defined.

Horty has lately changed his mind about reinstatement. He is now inclined to think, more in agreement with Prakken, that the problems with reinstatement are due indeed to how the information is represented.

> In the previous work, I relied on situations such as the [Example 3] to argue against reinstatement. I now believe, however, that the problem lies, not with reinstatement itself, but with our formalization of these situations – and that, once they are represented properly, reinstatement can be seen as innocuous [22]: 214–215).[10]

Of course, we agree about this diagnosis on reinstatement, and also recognize that some aspects of representation are determinant. But, on the other hand, we are trying to diminish the dependency on representation. In our opinion, solutions such as adding meta-level defaults should be avoided as much as possible. Instead, we are inclined to the use of general criteria of comparison among arguments – comparison that can be automated without necessity of added-by-users pieces of information like abnormality predicates, meta-rules, etc. In that sense, we think that our criterion of maximal specificity constitutes an advance in that direction.

Besides reinstatement, other principles have been introduced in order to validate the argumentation inferences of rule-based argumentation systems, mainly consistency and closure [8,9,13], which DeLP does not satisfy. These issues has been solved in ASPIC+ [31,27], instead. That seems to be an advantage with respect to DeLP with regards to our aim, thus it is sensible to think about using this system as an alternative approach to our proposal. On the other hand, possible adjustments on ASPIC+ to deal with our intuition on maximal specificity seem not to be trivial. Basically, the problem is that the system cannot deal with specificity unless it be explicitly represented by the appropriate introduction of either undercutting defeaters or preferences among rules. Anyway, we plan to explore this possibility in the future.

Finally, the reinstatement principle has been formally defined in three different versions in [1]: reinstatement, weak reinstatement and CF-reinstatement. All Dung's semantics satisfy all those versions, but other aspects have been discussed that seem to question the epistemological adequacy of Dung's argumentation theory [7].

6. Conclusion

The issue of reinstatement as a principle for argument systems was the subject of a serious criticism [21], while its defense (mainly that of [30]) has not been entirely satisfactory in our opinion. The criticism focuses only cases in which specificity is the comparison criterion among arguments. We argued here that the problem is that existing specificity based argument systems are not capable of effectively defeating all non-maximally specific arguments outrightly. We proposed a formal criterion of maximal specificity which, in accordance with early researches about inductive explanations [19], takes into account the total evidence represented in the knowledge base to avoid some implausible conclusions. We have argued that specificity is not only a criterion for solving inconsistencies – condition (C1) – but also for eschewing some imprudent, exaggerated, magnified, or simply unsound conclusions – condition (C2). This last condition led us to formally introduce the notion of *cautiousness*, from which we also extracted the notion of *cautious defeater*. Moreover, we showed formal properties of these notions and proposed a procedure for finding cautious defeaters.

On the other hand, a dialectical approach to warrant was used to show that the rejection of non-maximally specific arguments and their undesired interferences can be captured by defining specific game protocols. We defined two protocols, Protocol 1 and Protocol 2, for skeptical and credulous behavior, respectively, and showed full correspondences between the winning strategies of the proponent and the warranted arguments sanctioned by Dung's grounded (under Protocol 1) and preferred (under Protocol 2) semantics.

Finally, we insist that our notion of maximal specificity was not fully characterized, since only necessary conditions were identified. Though we have not found any other exceptions to maximal specificity, neither in the defeasible reasoning literature nor in our own research, our proposal is open to further studies in order to get not only necessary but also sufficient conditions for a full characterization. Furthermore, an implementation of our theory is planned as future work.

Acknowledgments

We thank two anonymous reviewers who made very detailed and helpful criticisms that improved this paper. We also thank Rodrigo Iglesias for providing important insights on section 3.4.

Funding: This work was supported by Agencia Nacional de Promoción Científica y Técnica [PICT 2013-1489]; Universidad Nacional del Sur [PGI 24/I223]; and Universidad Católica de Cuyo (Argentina).

References

[1] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, Artif. Intell. 171 (10–15) (2007) 675–700.

[10] This quotation was adverted by a referee.

[2] P. Baroni, M. Giacomin, Semantics of abstract argument systems, in: I. Rahwan, G.R. Simari (Eds.), Argumentation in Artificial Intelligence, Springer Publishing Company, 2009, pp. 25–44.
[3] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artif. Intell. 128 (2001) 203–235.
[4] P. Besnard, A. Hunter, Elements of Argumentation, The MIT Press, 2008.
[5] G. Bodanza, C. Alessio, Reinstatement and the requirement of maximal specificity in argument systems, in: U. Kohlenbach, P. Barceló, R. de Queiroz (Eds.), Logic, Language, Information and Computation. Proceedings of the 21st International Workshop, WoLLIC, Valparaíso, Chile, September 1–4, 2014, in: Lect. Notes Comput. Sci., vol. 8652, Springer-Verlag, Berlin, Heidelberg, 2014, pp. 81–93.
[6] G. Bodanza, F. Thomé, G. Simari, Argumentation games for admissibility and cogency criteria, in: B. Verheij, S. Szeider, S. Woltran (Eds.), Computational Models of Argument, in: Proceedings of COMMA 2012, IOS Press, 2012, pp. 153–164.
[7] G. Betz, Assessing the epistemological relevance of dung-style argumentation theories, Ann. Math. Artif. Intell. (2015), http://dx.doi.org/10.1007/s10472-015-9484-4, published online.
[8] M. Caminada, L. Amgoud, An axiomatic account of formal argumentation, in: Proceedings of the 20th National Conference on Artificial Intelligence (AAAI 2005), 2005, pp. 608–613.
[9] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, Artif. Intell. 171 (5–6) (2007) 286–310.
[10] E. Dijkstra, A Discipline of Programming, Prentice Hall, NJ, 1976.
[11] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–358.
[12] P.M. Dung, T.C. Son, An argument-based approach to reasoning with specificity, Artif. Intell. 133 (2001) 35–85.
[13] P.M. Dung, P.M. Thang, Closure and consistency in logic-associated argumentation, J. Artif. Intell. Res. 49 (2014) 79–109.
[14] D. Etherington, Formalizing nonmonotonic reasoning systems, Artif. Intell. 31 (1987) 41–85.
[15] D. Etherington, R. Reiter, On inheritance hierarchies with exceptions, in: Proceedings of the Third National Conference on Artificial Intelligence (AAAI-83), 1983, pp. 104–108.
[16] S.E. Fahlman, NETL: A System for Representing and Using Real-World Knowledge, MIT Press, Cambridge, MA, 1979.
[17] J. Fetzer, Carl Hempel, in: R. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2014 Edition), http://plato.stanford.edu/archives/fall2014/entries/hempel/.
[18] A. García, G. Simari, Defeasible logic programming: an argumentative approach, Theory Pract. Log. Program. 4 (1) (2004) 95–138.
[19] C. Hempel, Maximal specificity and lawlikeness in probabilistic explanation, Philos. Sci. 35 (2) (1968) 116–133.
[20] J. Horty, Some direct theories of nonmonotonic inheritance, in: D. Gabbay, C. Hobber, J. Robinson (Eds.), Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3: Nonmonotonic Reasoning and Uncertain Reasoning, Oxford University Press, 1994, pp. 111–187.
[21] J. Horty, Argument construction and reinstatement in logics for defeasible reasoning, Artif. Intell. Law 9 (2001) 1–28.
[22] J. Horty, Reasons as Defaults, Oxford University Press, 2012.
[23] J. Horty, R. Thomason, D. Touretzky, A skeptical theory of inheritance in nonmonotonic semantic networks, Artif. Intell. 42 (1990) 311–348.
[24] R. Loui, Defeat among arguments: a system of defeasible inference, Comput. Intell. 3 (1) (1987) 100–106.
[25] S. Modgil, Value based argumentation in hierarchical argumentation frameworks, in: P. Dunne, T. Bench-Capon (Eds.), Proc. of Computational Models of Argument, COMMA 2006, Liverpool, UK, September 11–12, 2006, in: Front. Artif. Intell. Appl., vol. 144, IOS Press, 2006, pp. 297–308.
[26] S. Modgil, M. Caminada, Proof theories and algorithms for abstract argumentation frameworks, in: I. Rahwan, G. Simari (Eds.), Argumentation in AI, Springer-Verlag, 2009, pp. 105–132.
[27] S. Modgil, H. Prakken, The ASPIC+ framework for structured argumentation: a tutorial, Argument and Computation 5 (1) (2014) 31–62.
[28] J. Pollock, Defeasible reasoning with variable degrees of justification, Artif. Intell. 133 (2001) 233–282.
[29] D. Poole, On the comparison of theories: preferring the most specific explanation, in: Proc. of the Ninth IJCAI, Los Altos, 1985, pp. 144–147.
[30] H. Prakken, Intuitions and the modelling of defeasible reasoning: some case studies, in: S. Benferhat, E. Giunchiglia (Eds.), Proceedings of the 9th International Workshop on Non-Monotonic Reasoning (NMR 2002), Tolouse, France, April 19–21, 2002, in: NMR, vol. 2, 2002, pp. 91–102.
[31] H. Prakken, An abstract framework for argumentation with structured arguments, Argument and Computation 1 (2) (2011) 93–124.
[32] R. Reiter, G. Criscuolo, On interacting defaults, in: Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81), 1981, pp. 270–276.
[33] G. Simari, R. Loui, A mathematical treatment of defeasible reasoning, Artif. Intell. 53 (1992) 125–157.
[34] F. Stolzenburg, A. García, C. Chesñevar, G. Simari, Computing generalized specificity, J. Appl. Non-Class. Log. 13 (1) (2003) 87–113.
[35] S. Toulmin, The Uses of Argument (Updated edition first published), Cambridge University Press, 2003.
[36] G. Vreeswijk, H. Prakken, Credulous and sceptical argument games for preferred semantics, in: Proc. JELIA 2000, in: LNAI, vol. 1919, 2000, pp. 239–253.
[37] C. Wirth, F. Stolzenburg, A series of revisions of David Poole's specificity, Ann. Math. Artif. Intell. (2015) 1–54.