

## An appreciation of John Pollock's work on the computational study of argument

Henry Prakken<sup>a\*</sup> and John Horty<sup>b</sup>

<sup>a</sup>*Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands & Faculty of Law, University of Groningen, Groningen, The Netherlands;* <sup>b</sup>*Philosophy Department, Institute for Advanced Computer Studies and Computer Science Department, University of Maryland, College Park, MD, USA*

(Received 12 January 2012; final version received 19 January 2012)

John Pollock (1940–2009) was an influential American philosopher who made important contributions to various fields, including epistemology and cognitive science. In the last 25 years of his life, he also contributed to the computational study of defeasible reasoning and practical cognition in artificial intelligence. He developed one of the first formal systems for argumentation-based inference and he put many issues on the research agenda that are still relevant for the argumentation community today. This paper presents an appreciation of Pollock's work on defeasible reasoning and its relevance for the computational study of argument. In our opinion, Pollock deserves to be remembered as one of the founding fathers of the field of computational argument, while, moreover, his work contains important lessons for current research in this field, reminding us of the richness of its object of study.

**Keywords:** formal models of argumentation; non-monotonic accounts of argument; defeasible reasoning

### 1. Introduction

John Pollock (1940–2009) was an influential American philosopher who made important contributions to various fields, including epistemology and cognitive science. In the last 25 years of his life, he also contributed to artificial intelligence (AI), first to the study of defeasible reasoning and then to the study of decision-theoretic planning and practical cognition. In his work on defeasible reasoning, Pollock developed one of the first formal systems for argumentation-based inference and he put many issues on the research agenda that are still relevant for our community today. This paper reviews the relevance of Pollock's work on defeasible reasoning for the computational study of argumentation.<sup>1</sup> His later work on rational decision-making and practical cognition (including decision-theoretic planning) will not be discussed since, unlike his work on defeasible reasoning, it is not argumentation based.

There are many reasons to remember and acknowledge Pollock's work in this journal. Many important topics in our field were first studied by Pollock, or first studied in detail, such as argument structure, the nature of defeasible reasons, the interplay between deductive and defeasible reasons, rebutting versus undercutting defeat, argument strength, argument labellings, self-defeat, and resource-bounded argumentation. Another reason to remember Pollock is that, these days, several lessons to be learned from his work tend to be forgotten and several important issues that he studied tend to be neglected for the sake of technical simplicity but at the expense of cognitive adequacy – especially in two current research strands: namely, work on abstract argumentation and work

---

\*Corresponding author. Email: henry@cs.uu.nl

on classical and deductive argumentation. This is particularly unfortunate since a central virtue of an argumentation approach is its grounding in natural and intuitive concepts; if conceptual naturalness is sacrificed for technical simplicity, then our field is in danger of becoming sterile and inward looking and failing to realise its high potential. A secondary aim of this paper, therefore, is to use Pollock's legacy to remind the research community of the richness of its object of study.

A concise quote that summarises Pollock's view on defeasible reasoning is as follows:

Defeasible reasoning is, *a fortiori*, reasoning. Reasoning proceeds by constructing arguments, where *reasons* provide the atomic links in arguments. *Conclusive reasons* logically entail their conclusions. Defeasibility arises from the fact that not all reasons are conclusive. Those that are not are *prima facie reasons*. Prima facie reasons create a presumption in favour of their conclusion, but it is defeasible. (1995, p. 85)

Pollock thus depicts arguments as inference trees, where the nodes are statements, with the leaf nodes being premises, and the links are applications of 'reasons'. He thus regarded reasons as inference rules, but he did not identify inference rules with deductive inference rules alone. Pollock strongly emphasised the importance of *defeasible* reasons in argumentation.<sup>2</sup> He was quite insistent that defeasible reasoning is not just some exotic, exceptional, add-on to deductive reasoning – or, as is sometimes thought in computer science, only a heuristic matter – but is, instead, an essential ingredient of our cognitive life:

It is supposed that defeasible reasoning is less secure than normal reasoning, and should be countenanced only for the sake of computational efficiency. Its use is not just a matter of computational efficiency. It is logically impossible to reason successfully about the world around us using only deductive reasoning. All interesting reasoning outside mathematics involves defeasible steps. (Pollock 1995, p.41)

... we cannot get around in the world just reasoning deductively from our prior beliefs together with new perceptual input. This is obvious when we look at the varieties of reasoning we actually employ. We tend to trust perception, assuming that things are the way they appear to us, even though we know that sometimes they are not. And we tend to assume that facts we have learned perceptually will remain true, as least for a while, when we are no longer perceiving them, but of course, they might not. And, importantly, we combine our individual observations inductively to form beliefs about both statistical and exceptionless generalizations. None of this reasoning is deductively valid. (Pollock 2009, p. 173)

Starting in the 1980s, Pollock set out to formalise this view of defeasible reasoning and then to implement it in an automated reasoner that he baptised as OSCAR. Besides giving a general account of the structure of arguments and of the interplay between deductive and defeasible inferences, he also formalised particular defeasible reasons that he found important in human cognition. In particular, he formalised reasons for perception, memory, induction, the statistical syllogism, and temporal persistence, as well as the so-called undercutting defeaters for these reasons. Pollock extensively studied the problem of identifying the justified beliefs generated by a set of arguments and their defeat relations, and his various solutions to this problem predated much current work on argumentation-based semantics.

In what follows, we first give a historic sketch of Pollock's work and its relation with other work in philosophy and AI. We then review the essentials of his formal models, after which we critically examine some of the current work on deductive argumentation in light of Pollock's work. We end with some observations on Pollock's way of working and thinking and a summary of his contributions to our field.

## 2. A historic sketch

In modern philosophy, the study of defeasibility and defeat originated in legal philosophy, in the work of Hart (1949), who pointed out, first, how otherwise binding contracts might be compromised

by the presence of defeating conditions and later emphasised the defeasible nature of legal rules in general. The concept of defeasibility – echoing Ross’s notion of *prima facie* rules Ross (1930) as well as some of Wittgenstein’s ideas – was originally studied solely within legal philosophy, practical reasoning, and ethics. It made the jump from ethics to epistemology in the work of Chisholm (1957), who appealed to the idea in both fields, and later in the work of Pollock, beginning with his (1970) and developed in a number of articles leading up to his (1974) and beyond.

Pollock’s work on argumentation originated as an attempt to make formal sense of the intuitive notion of defeasible reasoning that seemed to be at work in these papers and books. In fact, the task had been attempted before. There is an early paper by Chisholm (1974), a heroic effort whose failure is no surprise given the limited tools available at the time. Still, in spite of the blossoming of philosophical logic in the 1960s and 1970s, the logical study of defeasible reasoning had received almost no attention at all. It is fair to say that Pollock, working in isolation, was the first philosopher working in the field of philosophy, as opposed to computer science, to outline an adequate framework for defeasible reasoning

Pollock’s initial paper on the topic was his classic (1987). By the time that paper was published, several researchers in AI had independently begun to explore an argument-based approach to defeasible, or non-monotonic, reasoning. This early research includes the work of Touretzky (1986) on inheritance systems, later developed along with several collaborators (Horty, Thomason, and Touretzky 1987, 1990; Touretzky, Horty, and Thomason 1987), as well as the work of Nute (1988) and Loui (1987). And of course, by the late 1980s, the field of non-monotonic reasoning more generally, of which the argument-based approach was only a part, had been recognised as an important subfield of AI. Although Pollock’s ideas originated in his efforts to understand defeasible reasoning in a philosophical context, it was the formation, within AI, of a community of researchers focused on non-monotonic reasoning that led to the publication of these ideas. Concerning his 1987 paper, Pollock later wrote that he first developed the idea in 1979, but that he did not initially publish it because, as he says, ‘being ignorant of AI, I did not think anyone would be interested’ (Pollock 2007b, p. 469). It is interesting to note that if Pollock had published this idea when it first occurred to him, the result would have been not only the first argument-based theory of defeasible reasoning, but also one of the first systems of any kind for non-monotonic reasoning.

In any case, the paper was eventually published and then several successive papers or books on argumentation and defeasibility (Pollock 1992, 1994, 1995, 2002, 2007a,b, 2009, 2010), and Pollock began a fruitful period of interaction with researchers in non-monotonic reasoning, AI more generally, and of course, argumentation. Although Pollock learned much from these communities, they also learned much from him. Research in non-monotonic reasoning, at the time, was motivated by a set of concerns from planning, logic programming, knowledge representation, and database theory. Pollock brought a fresh set of concerns, from his earlier work in traditional epistemology, along with some fresh ideas. The most important of these may have been the distinction, introduced by Pollock (1970) between two separate kinds of defeat: rebutting and undercutting. In fact, something like this distinction had emerged independently in the field of knowledge representation, in the ‘uncancel’ links from Fahlman’s (1979) semantic networks; but the idea quickly evaporated in the formal treatments of these networks, for reasons of theoretical simplicity. It was Pollock’s insistence on the importance of this distinction, for reasons of descriptive adequacy, that reintroduced it into the fields of non-monotonic reasoning and argumentation, where it has remained vitally important, for example, in representation problems involving legal reasoning. Indeed, Pollock’s idea of an undercutting defeater is closely related to the notion of an exclusionary reason, first introduced by Raz (1975), who himself cites Pollock; for more recent work linking the two ideas, see Horty (2012).

### 3. A closer look at Pollock's work on defeasible reasoning

Let us now take a closer look at Pollock's approach to modelling defeasible reasoning. His first paper with a formal system for argumentation-based defeasible reasoning was his (1987), but he published several later versions, notably his (1992, 1994, 1995, 2002, 2009), largely because he changed his mind on several design issues, especially on the characterisation of defeasible inference, and on the notion of argument strength.

#### 3.1. Constant features

There are several constant features in Pollock's formalisation of defeasible reasoning. Reasoning proceeds from a knowledge base of classical-logic formulas by chaining reasons into inference trees, where all reasons are either deductive or defeasible. Only applications of defeasible reasons can be defeated, and there are two kinds of defeaters: *rebutting* defeaters attack the conclusion of a defeasible inference by favouring a conflicting conclusion, while *undercutting* defeaters attack the defeasible inference itself, without favouring a conflicting conclusion. The concept of undercutting can be illustrated with Pollock's own favourite example: if the object looks red, that is a reason for concluding, defeasibly, that the object is red; but the presence of red illumination interrupts the reason relation without suggesting any conflicting conclusion. In later iterations of Pollock's systems, inferences and conclusions have probabilistic strengths, and these strengths partly determine defeat, in that an attempted defeater only succeeds if it is not weaker than its target.

Pollock's technical presentation of his system differs with his different publications, but the basic ideas can be sketched as follows. Technically, Pollock considers sequences of *lines* from an *argument* (in later work, he speaks of *nodes* from an *inference graph*). Each line from an argument is a tuple  $(\varphi, r, l, s)$ , where  $\varphi$  is a proposition,  $r$  is the reason applied to infer  $\varphi$  (where this reason can also be that  $\varphi$  is taken from the knowledge base),  $l$  is the set of preceding lines from which  $\varphi$  is inferred, and  $s$  is the line's strength. A sequence of such lines is a (linear) *argument* if each line is such that its proposition is either inferred from earlier lines or taken from the knowledge base. Thus, Pollock's notion of an argument is similar to the familiar notion of a deduction. In a suppositional argument, lines also have a set of suppositions; these can be added to each line and can also be used to infer conclusions; their retraction gives rise to conditional conclusions, just as in natural deduction. In this paper, we will only consider linear arguments.

The *defeat* relation first among argument lines and then among arguments can then be defined as follows:

DEFINITION 3.1 (Defeat among argument lines and arguments)

- (1) An argument line  $(\varphi, r, l, s)$  *defeats* an argument line  $(\varphi', r', l', s')$  iff
  - (a)  $r'$  is a defeasible rule,
  - (b)  $s \geq s'$ , and
  - (c)  $\varphi = \neg\varphi'$  or  $\varphi = \neg r'$  (here  $\neg r$  is shorthand for saying that the antecedents of rule  $r$  do not support its consequent).
- (2) An argument  $A$  *defeats* an argument  $B$  iff a line of  $A$  defeats a line of  $B$ .

Consider by way of example the following (informal) version of the well-known Tweety example, with the arguments displayed in tree form (as is well known, each deduction can be converted into an inference tree, while each inference tree can be converted into several deductions, each capturing an order in which a reasoner can construct the tree). Figure 1 shows three (maximal) arguments: two arguments for the conclusions that Tweety flies, respectively, does not fly, and an argument denying that the reason used to infer the first of these conclusions applies.

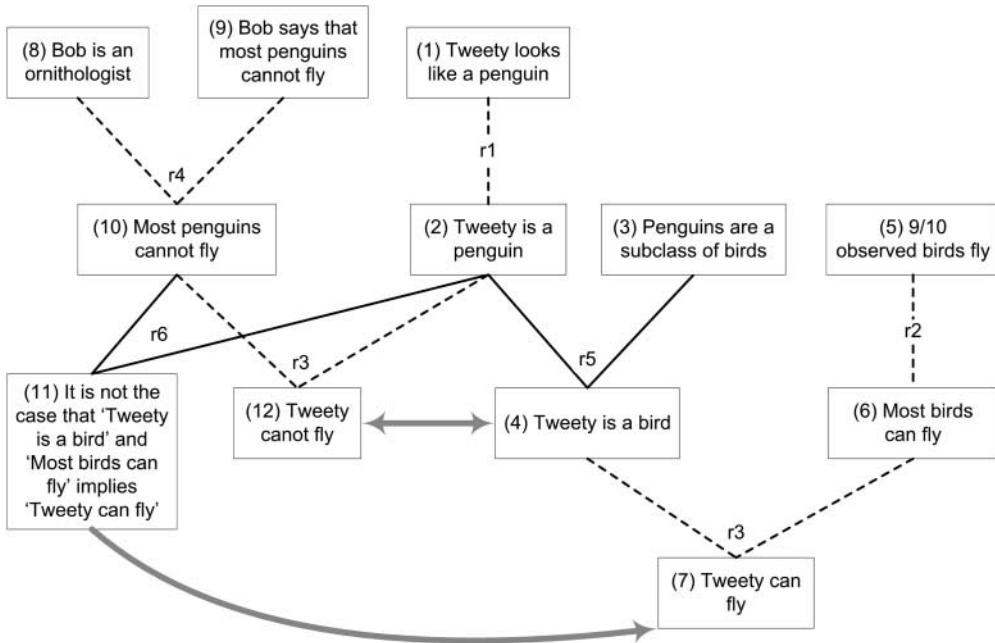


Figure 1. An example.

Figure 1 assumes four defeasible inference rules, informally paraphrased as follows:

- $r_1$ : That an object looks like having property  $P$  is a defeasible reason for believing that the object has property  $P$
- $r_2$ : That  $n/m$  observed  $P$ 's are  $Q$ 's (where  $n/m > 0,5$ ) is a defeasible reason for believing that most  $P$ 's are  $Q$ 's
- $r_3$ : That most  $P$ 's are  $Q$ 's and  $x$  is a  $P$  is a defeasible reason for believing that  $x$  is a  $Q$
- $r_4$ : That an ornithologist says  $\varphi$  about birds is a defeasible reason for believing  $\varphi$

Rule  $r_1$  expresses that perceptions yield a defeasible reason for believing that what is perceived to be the case is indeed the case, and rule  $r_2$  captures enumerative induction, while  $r_3$  expresses the statistical syllogism. Rule  $r_4$  can be seen as a special case of the argumentation scheme from expert testimony; cf. Walton (1996) (this is our way of illustrating that Pollock's notion of prima reasons is very similar to Walton's notion of an argumentation scheme; Pollock would probably have depicted this inference as an application of the statistical syllogism to the generalisation 'most experts speak the truth about their field of expertise').

Moreover, Figure 1 assumes an obvious strict inference rule plus an undercutting defeater for  $r_3$ :

- $r_5$ : That  $P$ 's are a subclass of  $Q$ 's and  $a$  is a  $P$  is a deductive reason for believing that  $a$  is a  $Q$
- $r_6$ : That  $x$  is an  $R$ , most  $R$ 's are not  $Q$ 's and  $R$ 's are a subclass of  $P$ 's is a deductive reason for believing  $\neg r_3$

Rule  $r_6$  is a special case of Pollock's 'subproperty defeater' of the statistical syllogism, which says that conflicting statistical information about a subclass undercuts the statistical syllogism for the superclass.

With these inference rules made explicit, the three arguments can be formally represented as follows (ignoring strength, so that argument lines can be depicted as triples rather than as four-tuples):

A:

- 1: (Tweety looks like a penguin, fact,  $\emptyset$ )
- 2: (Tweety is a penguin,  $r_1$ , {1})
- 3: (Penguins are a subclass of birds, fact,  $\emptyset$ )
- 4: (Tweety is a bird,  $r_5$ , {2, 3})
- 5: ( $\frac{9}{10}$  observed birds fly, fact,  $\emptyset$ )
- 6: (Most birds can fly,  $r_2$ , {5})
- 7: (Tweety can fly,  $r_3$ , {4, 6})

B:

- 1: (Tweety looks like a penguin, fact,  $\emptyset$ )
- 2: (Tweety is a penguin,  $r_1$ , {1})
- 3: (Penguins are a subclass of birds, fact,  $\emptyset$ )
- 8: (Bob is an ornithologist, fact,  $\emptyset$ )
- 9: (Bob says that most penguins cannot fly, fact,  $\emptyset$ )
- 10: (Most penguins cannot fly,  $r_4$ , {8, 9})
- 11: ( $\neg r_3$ ,  $r_6$ , {2, 3, 10})

C:

- 1: (Tweety looks like a penguin, fact,  $\emptyset$ )
- 2: (Tweety is a penguin,  $r_1$ , {1})
- 8: (Bob is an ornithologist, fact,  $\emptyset$ )
- 9: (Bob says that most penguins cannot fly, fact,  $\emptyset$ )
- 10: (Most penguins cannot fly,  $r_4$ , {8, 9})
- 12: (Tweety cannot fly,  $r_3$ , {2, 10})

Note that these arguments have several ‘subarguments’, since any sequence of lines of which all elements are a fact or inferred from previous elements in the sequence is an argument.<sup>3</sup>

In Figure 1, deductive, respectively, defeasible, inferences are visualised with, respectively, solid and dotted lines without arrow heads, while defeat relations are displayed with arrows. Naming arguments according to their last line, it can be seen that argument  $B_{11}$  strictly defeats argument  $A_7$  since line 11 undercuts line 7, while arguments  $A_7$  and  $C_{12}$  defeat each other since lines 7 and 12 rebut each other.

As should be apparent from this example, Pollock’s notion of an argument as a tree or sequence of inferences is quite natural; its only remarkable features are that arguments interleave application of deductive and defeasible inference rules and that arguments can have varying strengths. His notion of defeat is also quite natural. First, given that deductive reasons provide conclusive support for their conclusion, it is natural that their application cannot be attacked: one cannot at the same time rationally accept the premises and deny the conclusion of a deductive inference. Second, Pollock’s distinction between rebutting and undercutting defeat, while new when he introduced it in 1970, has meanwhile proven its value.<sup>4</sup>

### 3.2. Semantics

While, throughout his career, Pollock left his notions of argument construction and defeat essentially unchanged, he more than once changed the semantics for his system. For Pollock, a *semantics* was an account of how the set of constructed arguments, taken together with their defeat relations, determines what a cogniser should believe. For today’s students of computational argument, the

approach to follow might seem to be obvious: since Pollock's system results in a set of arguments together with a binary relation of defeat, he could simply have appealed to any of Dung's (1995) semantics of abstract argumentation frameworks. However, much of Pollock's work was published before Dung's seminal paper, and Pollock never explicitly used Dung's semantics. Nevertheless, his first two proposals have each been proven to be equivalent to one of Dung's proposals.

In his initial paper (Pollock 1987), after specifying an argument as *self-defeating* if one of its lines defeats another of its lines, Pollock defined his semantics by introducing the concepts of arguments that are in or out at various levels, and then ultimately undefeated, as follows:

DEFINITION 3.2 (Semantics of Pollock 1987)

- (1) All arguments that are not self-defeating arguments are *in at level 0*.
- (2) An argument is *in at level  $n + 1$*  iff it is not defeated by any argument in at level  $n$ .
- (3) An argument is *ultimately undefeated* iff there is an  $m$  such that for every  $n \geq m$ , the argument is in at level  $n$ .

In the example shown in Figure 1, all displayed arguments are *in* at level 0. The only arguments that are not *in* at level 1 are  $B_7$  and  $C_{12}$ . Of these, only  $C_{12}$  is *in* at level 2, since its only defeater, which is  $A_7$ , is not *in* at level 1. Furthermore, all other arguments, including  $B_{11}$ , remain *in* at all levels, so  $A_7$  remains not *in* at all levels; hence all arguments except  $A_7$  are ultimately undefeated.

Dung (1995) proved that if there are no self-defeating arguments, then Definition 3.1 is equivalent to his grounded semantics. More precisely, Dung first observed that the ideas from Definition 3.1 can be defined in terms of an operator that, for a given set of arguments, returns the set of arguments undefeated by any argument in that set. Dung then proved that his 'characteristic function' of argumentation frameworks, which for a given set of arguments returns all arguments all of whose defeaters are themselves defeated by some argument from that set, can be defined as a double application of this operator. Finally, Dung proved that Pollock's own level construction yields the least fixed point of this characteristic function, leading to what Dung calls the grounded extension.

After his initial semantics, Pollock later (1994, 1995) turned to a labelling-based approach, which, moreover, does not refer to arguments but instead relies on the notion of an *inference graph*: nodes in such a graph correspond to lines of argument and links represent either reason or defeat relations. An example of such an inference graph is that shown in Figure 1. Despite his new focus on inference graphs, rather than on sets of arguments, Pollock continued to present his work as argumentation based:

The theory of defeasible reasoning adumbrated in this book is an 'argument-based' theory, in the sense that it characterizes defeasible consequence in terms of the interactions between the inference steps of all possible arguments that can be constructed from a given set `input` using a fixed set of defeasible reasons and defeaters. (Pollock 1995, p. 105)

Pollock's new definition moves through the idea of a partial defeat status assignment, which again labels nodes as in or out, to that of a maximal defeat status assignment, which is as complete a status assignment as possible. Arguments are then characterised as ultimately undefeated, defeated outright, or provisionally defeated depending upon their behaviour in the various maximal status assignments.

DEFINITION 3.3 (Semantics of Pollock 1994, 1995) An assignment  $\sigma$  of *in* and *out* to a subset of the nodes of an inference graph is a *partial defeat status assignment* iff

- (1)  $\sigma$  assigns *in* to any initial node;

- (2)  $\sigma$  assigns *in* to a non-initial node  $\alpha$  iff  $\sigma$  assigns *in* to all immediate ancestors of  $\alpha$  and  $\sigma$  assigns *out* to all nodes defeating  $\alpha$ ; and
- (3)  $\sigma$  assigns *out* to a node  $\alpha$  iff  $\sigma$  either assigns *out* to an immediate ancestor of  $\alpha$  or assigns *in* to a node defeating  $\alpha$ .

A *defeat status assignment* is a maximal partial defeat status assignment; a node is *ultimately undefeated* if it is *in* in all defeat status assignments, *defeated outright* if it is *in* in no defeat status assignments and *provisionally defeated* otherwise.

In the example shown in Figure 1, lines 1, 3, 5, 8, and 9 must be assigned *in* by clause (1). Then, all lines except lines 7 and 12 must be assigned *in* by clause (2) since they have no defeaters. Then, line 7 must be assigned *out* by clause (3), since it is defeated by line 11, which must be assigned *in*. Then, line 12 must be assigned *in* by clause (2) since its only defeater must be assigned *out*.

Jakobovits (Jakobovits and Vermeir 1999; Jakobovits 2000) proved Definition 3.3 to be equivalent to another of Dung's proposals, the preferred semantics. Indeed, it is easy to see that the conditions on defeaters in (2) and (3) are the same as the conditions of preferred labellings in the sense of Caminada (2006). Such labellings label arguments of a Dung-style abstract argumentation framework as *in*, *out*, or *undecided*, in a way that satisfies the following constraints:

1. An argument is *in* if all arguments defeating it are *out*.
2. An argument is *out* if it is defeated by an argument that is *in*.
3. An argument is *undecided* otherwise.

A labelling is *preferred* if it maximises the set of *in* arguments, while it is *grounded* if it minimises this set. Jakobovits's result indicates that Pollock could also have formulated his new semantics by retaining his old notions of argument and defeat and directly using the above definition of preferred labellings. In any case, we can say that the idea of argument labellings was introduced in our field by Pollock, although its use in non-monotonic logic ultimately goes back to Doyle's (1979) justification-based truth maintenance systems.

The main reason for Pollock changing his mind between 1987 and 1994 was to refine his treatment of self-defeating arguments. While in his (1987) semantics, they were all unable to affect the status of other arguments, Pollock later realised that there are two kinds of self-defeating arguments, one of which should still be capable of preventing other arguments from being ultimately undefeated. The first kind of self-defeat he considered results in a situation in which two arguments for contradictory conclusions rebut each other (see Figure 2; in the examples given below, we assume for simplicity that all arguments are of equal strength). Here,  $r_1$  says that  $q$  is a defeasible reason for  $p$ , while  $r_2$  says that  $r$  is a defeasible reason for  $\neg p$ . Given  $q$  and  $r$  as facts, this results in two arguments rebutting each other. The contradictory conclusions can then be combined by applying a strict rule expressing the *Ex Falso* principle (an inconsistent set of formulas deductively implies everything) to support any formula. Thus, a rebuttal can be constructed for any other defeasible argument line, such as line 7 here. Clearly, such self-defeating rebuttals should not prevent any other argument from being ultimately undefeated, and Definition 3.1 respects this by declaring all self-defeating arguments as not *in* at level 0, so that they do not interfere with other arguments.

However, there is a second kind of self-defeating argument, which should be able to prevent other arguments from being justified. Consider the following version of the argument scheme from witness testimony plus an undercutter in case the witness is incredible:

- $r_1$ : That a witness says  $\varphi$  is a defeasible reason for  $\varphi$
- $r_2$ : That a witness is incredible is a deductive reason for  $\neg r_1$



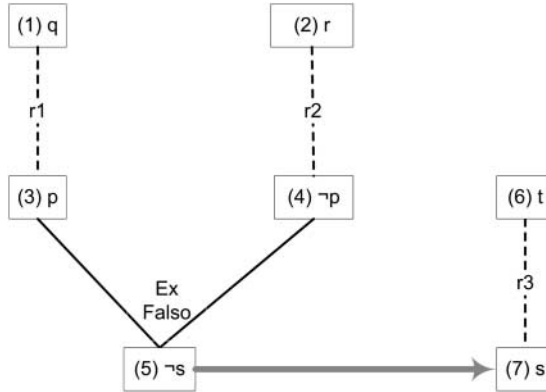


Figure 2. Parallel self-defeat.

Assume as given that Witness John says that he is incredible. Then (again ignoring strength), we can construct the following argument (on the left of Figure 3):

- 1: (Witness John says he is incredible, fact,  $\emptyset$ )
- 2: (John is incredible,  $r_1$ , {1})
- 3: ( $\neg r_1$ ,  $r_2$ , {2})

The argument up to line 3 is self-defeating, since line 3 undercuts line 2. Thus, according to Definition 3.1, the argument up to line 3 is at no level *in*. But then the argument up to line 2 is ultimately undefeated since its only defearer is at no level *in*, while yet a deductive consequence of the conclusion of line 2 (the conclusion of line 3) cannot be drawn, since the argument up to line 3 is not ultimately undefeated. This is strange.<sup>5</sup> By contrast, according to Definition 3.3, there is a unique preferred status assignment for this example, in which line 1 is in and both lines 2 and 3 are undefined. Thus, although both lines 2 and 3 are defeated outright, line 3 still retains its ability to prevent other argument lines from being ultimately undefeated, and this is desirable. Suppose witness John also says something completely unrelated, such as ‘The suspect hit the victim.’ We then also have the following argument (on the right of Figure 3):

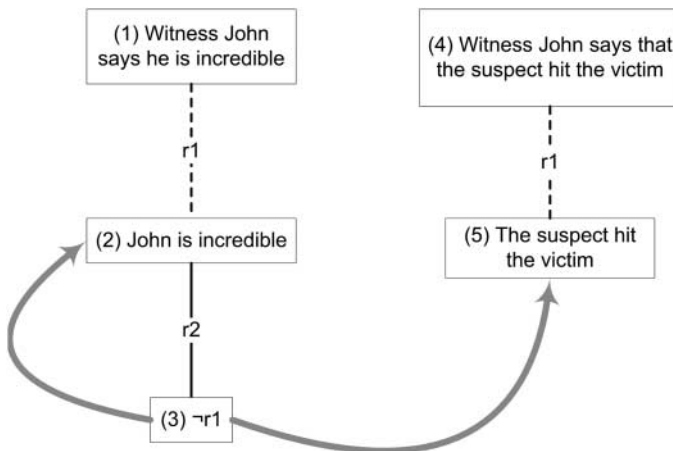


Figure 3. Serial self-defeat.

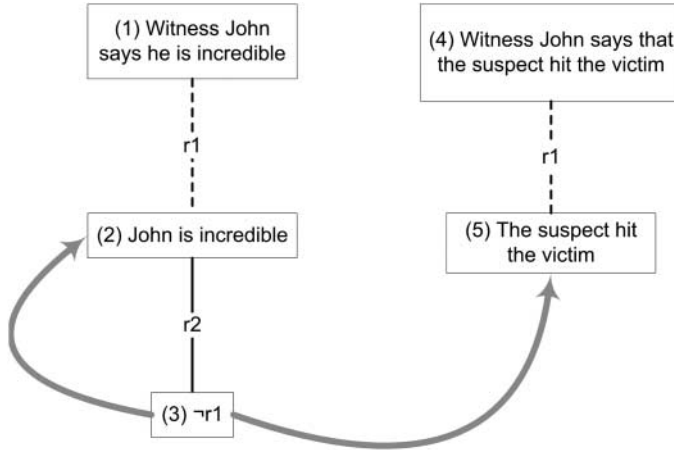


Figure 4. A self-defeating witness.

- 4: (Witness John says that the suspect stabbed the victim, fact),  $\emptyset$ )
- 5: (The suspect hit the victim,  $r_1$ , {4})

According to Definition 3.1, line 5 is ultimately undefeated, while yet it is based on a statement of a witness who says of himself that he is incredible. This seems counterintuitive. By contrast, according to Definition 3.3, line 5 is also defeated outright, since its status is undefined.<sup>6</sup>

In conclusion, Definition 3.3 captures that there are two classes of self-defeating arguments, which are both always defeated outright, but one of which still has the power of preventing other arguments from being ultimately undefeated. This observation also implies that self-defeating arguments cannot simply be ruled out from consideration by definition.

In one of his last publications (2009), Pollock again revised his semantics, motivated by concerns similar to those of Baroni and Giacomini (2005). He realised that his (1994) and (1995) distinction between two types of self-defeat in fact also gave different treatments to odd and even defeat cycles of arbitrary length, and he regarded that as counterintuitive. Let us extend the above example by replacing the single self-defeating witness with first an even defeat cycle of two witnesses and then an odd defeat cycle of three witnesses (Figure 5). Suppose first we have two witnesses Albert and Bob, who say of each other that they are unreliable, thus undercutting each other. Imagine also that Albert and Donald rebut each other on the issue whether the suspect hit the victim. We then have two maximal labellings: in one, lines 3 and 8 are in, while lines 6 and 10 are out, so both lines 8 and 10 are provisionally defeated and we can believe neither Bob nor Donald on whether the suspect hit the victim. However, consider next three witnesses Albert, Bob, and Carole, where Albert says that Bob is unreliable, Bob says that Carole is unreliable, and Carole says that Albert is unreliable, while Albert and Donald still rebut each other on whether the suspect hit the victim (Figure 6). Then, there is only one maximal labelling, in which line 10 is in while line 8 is out (and all of 3, 6, and 13 undecided). We cannot create a labelling in which we believe both Albert and Carole but not Bob or Donald, since Albert and Carole are involved in an odd defeat cycle. So, this yields that line 10 is ultimately undefeated while line 8 is defeated outright, so we can believe Donald that the suspect did not hit the victim. Thus, the justification status of Donald’s testimony depends on whether its attacker, Albert, is involved in an odd or an even defeat cycle. Baroni and Giacomini (2005) and Pollock (2009) regard this as counterintuitive.<sup>7</sup>

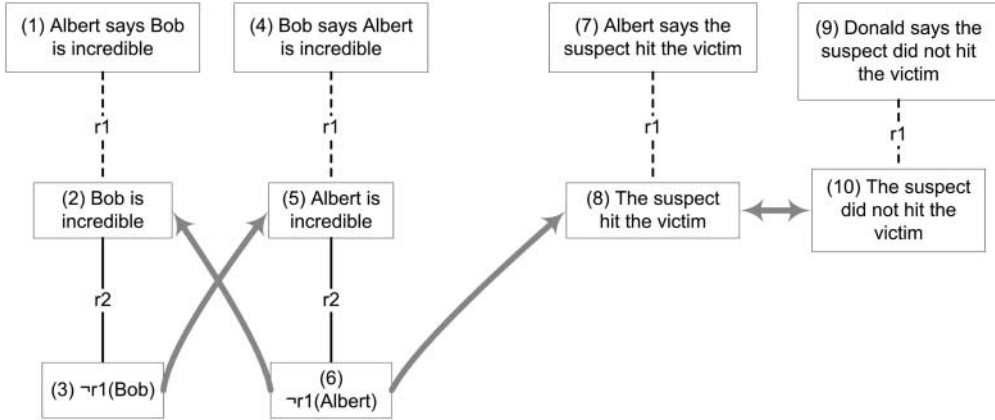


Figure 5. An even defeat cycle.

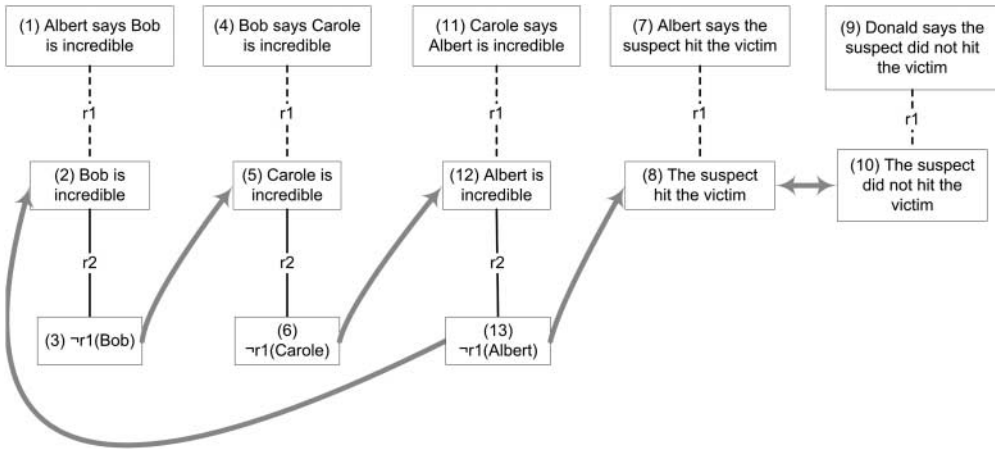


Figure 6. An odd defeat cycle.

### 3.3. Argument strength

Although Pollock’s earliest system, from 1987, did not yet include a notion of strength, Pollock later took the notion of strength of arguments very seriously. Since his system was meant for epistemic reasoning, he always formulated strength in terms of numerical degrees of belief. He was interested in computing the degree of justification of a statement  $P$ , that is, the degree of belief in  $P$  that an agent rationally ought to have. His approach here was non-standard. Against Bayesian approaches, he argued that degrees of belief and justification do not conform to the laws of probability theory. One argument he gave for this is that according to probability theory, necessary truths have probability 1, but if this is a degree of justification, then we would be equally justified in believing Fermat’s conjecture before and after Andrew Wiles proved it.

In his papers published in 1994 and 1995 (Pollock 1994, 1995), Pollock used a weakest-link approach to compute the strength of arguments: the strength of each conclusion is the minimum of the strengths of the inference with which it was derived and of the premises or intermediate conclusions from which it was derived. While these arguments can have various strengths, defeat is still an all-or-nothing matter in that defeaters that are weaker than their target cannot

affect the status of their target at all. In consequence, the justification status of a proposition for which arguments can be constructed is three valued: arguments can be ultimately defeated, ultimately undefeated, or provisionally defeated. However, in his (2002, 2007a, 2010), Pollock explored the idea that weaker defeaters can still weaken the justification status of their stronger targets. To formalise this, he now made the justification status of statements a matter of numerical degree, being a function of the strengths of both supporting and defeating arguments. In fact, Pollock seemed not fully sure that his 2002 account was the right one, witness the following quote:

In my (1995) I extended the above semantics to deal with reason strengths, but I am now convinced that the (1995) proposal was not correct. I tried again in my (2002), and that semantics or a minor variation of it may be correct, but I have not yet implemented it in OSCAR. (Pollock 2007b, p. 459)

Nevertheless, the basic idea that epistemic beliefs can be justified in varying degrees is very natural, and Pollock was right that the relation between defeasible reasoning and differing degrees of belief deserves more attention than it receives. His own proposal provides a good basis for further work on this topic.

### 3.4. *Defeasible reasons and argumentation schemes*

An important distinguishing feature of Pollock's account of defeasible inference rules is that he meant them to be general patterns of reasoning. While in AI there is a tradition to let defeasible inference rules express domain-specific information, as in, for example, default logic of Reiter (1980), Pollock's defeasible reasons are general patterns of epistemic defeasible reasoning. In particular, he formalised reasons for perception, memory, induction, temporal persistence and the statistical syllogism, as well as undercutters for these reasons. Here is how he contrasted his work with default logic:

In spirit, the theory of defeasible reasoning seems close to Reiter's default logic [34], with *prima facie* reasons and defeaters corresponding to Reiter's defaults. But there are also profound differences between the two theories. First, *prima facie* reasons are supposed to be logical relationships between concepts. It is a necessary feature of the concept red that something's looking red to me gives me a *prima facie* reason for thinking it is red. (To suppose we have to discover such connections inductively leads to an infinite regress, because we must rely upon perceptual judgments to collect the data for an inductive generalization.) By contrast, Reiter's defaults often represent contingent generalizations. If we know that most birds can fly, then the inference from being a bird to flying may be adopted as a default. In the theory of defeasible reasoning, the latter inference is instead handled in terms of the following *prima facie* reason schema: Most A's are B's, and this is an A is a *prima facie* reason for B . . . This is the statistical syllogism . . . (Pollock 1992, p. 9).

There is an interesting connection here with the literature of argumentation schemes (Walton, Reed, and Macagno 2008). Argumentation schemes are stereotypical non-deductive patterns of reasoning. Their use in building arguments is evaluated in terms of critical questions specific to a scheme. In the literature on argumentation theory, many collections of argumentation schemes have been proposed, for both epistemic and practical reasoning. Pollock's defeasible inference rules can in fact be seen as formalisations of some epistemic argumentation schemes. This also suggests a way to formalise reasoning with argumentation schemes (Prakken 2010b): they can be seen as defeasible inference rules and critical questions can be regarded as pointers to counterarguments. Some critical questions challenge an argument's premise and therefore point to premise attacks, others point to undercutting attacks, while again other questions point to rebutting attacks. Pollock's emphasis on the general nature of defeasible reasons plus his distinction between rebutting and undercutting defeat provided a basis for a formal framework for modelling

reasoning with argumentation schemes; the *ASPIC* framework as presented in Prakken (2010a) adds premise attack to Pollock's rebutting and undercutting attack and thus arguably provides a full framework for modelling reasoning with argumentation schemes.

### 3.5. *Suppositional reasoning*

In his earlier work, Pollock extended his system with suppositional reasoning, by allowing sets of assumptions to be introduced into and retracted from lines of argument just as in natural deduction. This validates, for example, a defeasible derivation of the material implication  $p \supset q$  from the fact that  $p$  is a defeasible reason for  $q$ . In fact, this feature of his system has not been taken up by others and Pollock himself did not use it any more in his later work.

### 3.6. *Partial computation*

Pollock also addressed the issue of partial computation. To deal with the intractability of the full version of his system, he suggested several alternative notions of defeat status, all making his system non-monotonic not just in the set of input beliefs but also in the amount of computation. One of these notions, 'justification', simply computes defeat status relative to the inference graph computed at a certain moment. Pollock also developed an alternative notion of adequacy of algorithms for defeasible reasoning, given that to be tractable, they cannot be sound and complete with respect to the ideal, that is, with respect to the set of all arguments that can be computed. For the details, we refer the reader to (Pollock 1995, chap. 4).

## 4. A critique of current computational models of argument in light of Pollock's work

Much current formal and computational work on argumentation is on abstract argumentation, as introduced by Dung (1995). However, to be useful and realistic, abstract models must be combined with accounts of the structure of arguments and the nature of attack and defeat. While this should be obvious, it is less obvious what such accounts should be. While almost all early work on argumentation in AI made a distinction between deductive (or 'strict') and defeasible inference rules, currently there is a growing body of work that models argumentation as inconsistency handling in classical or, more generally, deductive logic. In this section, we shall argue that Pollock's work strongly suggests that deductive argumentation is of limited applicability and that many, if not most, forms of argumentation can only be naturally modelled by combining deductive and defeasible inference rules.

As we have seen above, Pollock strongly emphasised the importance of defeasible reasons in argumentation. According to him, any full theory of argumentation should give an account of the interplay between deductive and defeasible reasons. In the 1980s and early 1990s, this view was quite in agreement with most of the then current research on non-monotonic logic.<sup>8</sup> Default logic (Reiter 1980), still one of the most influential non-monotonic logics, added defeasible inference rules to the proof theory of classical logic. Several systems for inheritance with exceptions (Horty and Thomason 1988) combined strict and defeasible inheritance rules. Simari and Loui (1992) fully formalised Loui's (1987) initial ideas on argumentation with strict and defeasible inference rules. This work in turn led to the development of Defeasible Logic Programming (Garcia and Simari 2004). Lin and Shoham (1989) proposed the idea of abstract argumentation structures with strict and defeasible rules and showed how a number of existing non-monotonic logics could be reconstructed as such structures. Gerard Vreeswijk further developed these ideas in his abstract argumentation systems (Vreeswijk 1997). Nute (1994) published the first version of Defeasible Logic, which also combines strict and defeasible domain-specific inference rules. Finally, Prakken

and Sartor (1997) formalised an argumentation logic with strict and defeasible inference rules and defeasible priorities explicitly as an instance of Dung's (1995) abstract argumentation frameworks. Currently, the proponents of the *ASPIC* framework (Prakken 2010a, Modgil and Prakken 2011) try to unify and further develop this work into a general framework for structured argumentation with both strict and defeasible inference rules.

Nowadays, however, there is a growing body of work that models argumentation as inconsistency handling in either classical logic or some other standard deductive logic (Besnard and Hunter 2001, 2008; Amgoud and Cayrol 2002; Parsons, Wooldridge, and Amgoud 2003; Amgoud and Besnard 2009; Gorogiannis and Hunter 2011). In Pollock's terms, this work regards all reasons as deductive. Accordingly, in these approaches arguments can only be attacked on their premises. If such a reduction is possible, then there is no need for new logics but just for a proper way of modelling inconsistency handling in deductive logic, which, so it is said, has the advantage that it is well understood (Besnard and Hunter 2008, p. 16).

Pollock did not include premise attack in his work, since he was only interested in what can be defeasibly inferred from a consistent body of information. When arguments are constructed with defeasible reasons, they can be attacked even if all their premises are accepted, since the premises only presumptively support their conclusion: it is rationally possible to accept all premises of a defeasible inference but still not accept its conclusion (at least if there are good reasons for not accepting it). Here, the philosophical distinction between *plausible* and *defeasible* reasoning is relevant; see Rescher (1976, 1977) and Vreeswijk (1993, chap. 8). Following Rescher, Vreeswijk described plausible reasoning as valid deductive reasoning from an uncertain basis and defeasible reasoning as deductively invalid (but still rational) reasoning from a solid basis. In these terms, models of deductive argumentation formalise plausible reasoning, while Pollock modelled defeasible reasoning. The question then becomes: can defeasible reasoning be reduced to plausible reasoning?

This question is not new. The current attempts to model argumentation on the basis of ordinary deductive logic have their parallel in the history of non-monotonic logic, in which there have been several attempts to reduce non-monotonic reasoning to some kind of inconsistency handling in classical logic; see, for example, Israel (1980), Poole (1988), Brewka (1989), and Baker and Ginsberg (1989).<sup>9</sup>

Now whether such a reduction is possible or not, it should at least be clear that it is a somewhat unnatural way to model defeasible reasoning, since the very idea of defeasible inference rules is that it is rationally possible to accept all their premises but still deny their conclusion. Consider the following well-known example: it is given that Quakers are normally pacifists, that Republicans are normally not pacifists, and that Nixon was both a Quaker and a Republican. There is nothing inconsistent in these givens – indeed, it is natural to think that they are all true. The reason is that ‘If  $Q$  then normally  $P$ ’ and ‘ $Q$ ’, taken together, do not deductively imply ‘ $P$ ’, since things could be abnormal: Nixon could be an abnormal Quaker or an abnormal Republican. A defeasible reasoner therefore does not have to reject any of the givens. Instead, such a reasoner wants to assume whenever possible that things are normal, in order to jump to conclusions about Nixon in the absence of evidence to the contrary.

Typical reductions of defeasible reasoning to inconsistency handling express such default assumptions as additional premises with a lower status than the rest of a knowledge base and model attacks on a defeasible inference as an attack on such premises. However, these approaches have been criticised for producing counterintuitive results due to the use of the material implication, which is claimed to be logically too strong for representing defeasible conditionals; see, for example, (Brewka 1991; Ginsberg 1994). While a review of this discussion goes beyond the scope of this paper (see for more details Prakken forthcoming), we can at least conclude from the existence of this debate – together with the vast body of work on defeasible reasons in philosophy,

non-monotonic logic, and argumentation theory – that the study of defeasible reasons deserves a central place in the formal and computational study of argumentation. One of Pollock’s main contributions to our field is the first formal account of defeasible reasons that is both technically mature and philosophically grounded. What he did not address, however, was the integration of defeasible with plausible reasoning, since he left no room for premise attack. Such an integration is one aim of the current *ASPIC* framework (Prakken 2010a; Modgil and Prakken 2011), which combines Pollock’s work on defeasible reasons with the more recent work on deductive argumentation.

## 5. Working style

In this section, we present some brief observations on Pollock’s way of working and thinking.

A remarkable aspect of Pollock’s work, especially given that he was a philosopher, is that he always implemented his theories of defeasible reasoning. Moreover, while most AI researchers have teams of graduate students to do their coding, Pollock mostly wrote his own code, in Common Lisp.

It is sometimes said that Pollock’s formalism for defeasible reasoning is too complex, but we do not think that this criticism is entirely fair. The main reason for the complexity of Pollock’s work is that his primary aim was not to design elegant and simple formalisms but to formalise defeasible reasoning in its full complexity. Therefore, the option to oversimplify formalisms just to be understood or to score theorems was not open to him. Moreover, as discussed above, several relations have been established between his work and Dung (1995)’s influential work on abstract argumentation, so that the place of his formalisms in the spectrum of argumentation approaches is now quite well understood.

However, an admirable aspect of his writings is that Pollock was always exceptionally clear and explicit about the reasons for and against his design choices. This relates to an equally admirable aspect of his thinking, namely his willingness to keep re-thinking and re-thinking his approach to defeasible reasoning. When he saw what he recognised as an error, Pollock was always willing to re-think even the most fundamental aspects of his existing theories. This is, for example, true for his final paper Pollock (2010), where he rejected his (1994, 1995) semantics and proposed an alternative. It is hard not to admire his intellectual honesty and his willingness, even at this late stage, to reformulate fundamental ideas in the face of a perceived difficulty.

## 6. A summary and evaluation of Pollock’s contributions to the field of computational argument

In summary, Pollock’s main contributions to the formal and computational study of argument are as follows:

- He proposed one of the first non-monotonic logics with explicit notions of argument and defeat.
- He introduced the important and now familiar distinction between rebutting and undercutting defeat.
- He was the first in AI to regard defeasible reasons as general principles of reasoning. He thus grounded his formalism in his work on epistemology and laid the basis for formalising argumentation schemes.

- The grounding of his theory in his work on epistemology allowed him to show that a full model of reasoning must include defeasible reasons and that defeasible reasoning cannot in general be reduced to inconsistency handling in deductive logic.
- He was the first to use a labelling approach in the semantics of argumentation (although derived directly from Doyle 1979).
- He took self-defeating arguments more seriously than anyone else, showing that they cannot simply be ruled out by definition but that some self-defeating arguments can still prevent other arguments from being justified.
- He took argument strength seriously and raised the issue of modelling degrees of justification.

His work also has some limitations.

- Some aspects of his work have not survived, such as his work on suppositional reasoning and on resource-bounded reasoning.
- While Pollock took argument strength seriously, he did not explicitly distinguish between attack and defeat, which sometimes leads to confusion (this matter is fully discussed in Horty 2012). Currently, there is a trend to clearly separate attack and defeat, e.g. Bench-Capon (2003), Prakken (2010a), Amgoud and Vesic (2011), Modgil and Prakken (2011), which, among other things, allows for clean explicit modellings of arguments about relative argument strength (Modgil 2009).
- In his work on argumentation, Pollock only modelled epistemic reasoning. He fully ignored normative reasoning and modelled practical reasoning (in his words ‘rational decision-making and practical cognition (including decision-theoretic planning)’)<sup>10</sup> without argumentation concepts (Pollock 1998, 1999, 2005). This makes his work less relevant for current argumentation models of practical reasoning, which is an important current theme in our field. Among other things, he gave no argumentation-based account of how practical reasoning depends on epistemic reasoning. Also, his work on argument strength is only relevant for epistemic reasoning.
- A related limitation is that since Pollock was focused on probabilistic strength of epistemic arguments, he always assumed that defeasible reasons can be arranged in a linear order of strength and never thought of incomparable strengths or about defeasible reasoning about the strength of defeasible arguments.

In conclusion, we can say that, above all, Pollock deserves to be remembered as one of the founding fathers of our field. Moreover, despite some limitations and imperfections, his work has historically been very influential while it still contains some important lessons for current research. Most importantly, Pollock’s work reminds us of the richness of our object of study, sometimes ignored in current work on, for example, abstract or deductive argumentation.

## Notes

1. We will confine ourselves to argumentation-based inference, since Pollock never studied argumentation-based dialogue.
2. Pollock varied in his terminology: in his earlier papers, he exclusively spoke of ‘conclusive’ and ‘prima facie’ reasons, while later he also referred to ‘deductive’ and ‘defeasible’ reasons (and sometimes to ‘inference rules’ instead of ‘reasons’). We will speak of deductive and defeasible reasons/inference rules. By ‘deductive argumentation’, we mean argumentation where all arguments are built with deductive inference rules, and by ‘classical argumentation’, we mean the special case of deductive argumentation where the inference rules consist of all valid propositional or first-order inference rules.



3. In systems like the *ASPIC* framework of Prakken (2010a), the notion of a subargument is refined to sequences which in graph form are trees. For example, according to Pollock (1987), the sequence (1, 3) is an argument but not according to Prakken (2010a).
4. Although some terminological confusion has arisen since others (Krause, Ambler, Elvang-Gøransson, and Fox 1995, Besnard and Hunter 2001, 2008, Amgoud and Cayrol 2002) have used it for attack on a premise instead of on the application of a defeasible inference rule.
5. Indeed, it violates Caminada and Amgoud's (2007) rationality postulate of closure of argument extensions under deductive inference.
6. In fact, Caminada (2005) showed that this treatment of self-defeat is not yet optimal. Consider again two rebutting arguments combined with Ex Falso to an argument for any proposition. Pollock thought that always at least one of the rebutting subarguments would be *out* so that the Ex Falso argument would also be *out*. However, Caminada showed that if both rebutting arguments have self-defeating arguments of the second type as a subargument, then (if there are no other defeaters) they are both undefined, so the Ex Falso argument is also undefined and retains its power to prevent other arguments from being *in*.
7. Against this this, Bench-Capon, personal communication, has argued that odd and even defeat cycles are logically different. According to him, odd defeat cycles are paradoxes, while even defeat cycles are dilemmas. If he is right, then a different treatment of odd and even defeat cycles is justified and Pollock would not have needed to revise his (1994, 1995) semantics.
8. Although Pollock's study of reasons as *general* patterns of reasoning sets his work apart from most other work in this vein, which often uses defeasible reasons for expressing *domain-specific* regularities.
9. Assumption-based argumentation (Bondarenko, Dung, Kowalski, and Toni 1997; Dung, Kowalski, and Toni 2009) is similar but more general; on the one hand, it only allows for premise attack and, on the other hand, it does not commit to classical or deductive logic as the source of its inference rules.
10. <http://oscarhome.soc-sci.arizona.edu/ftp/publications.html>, accessed on 28 November 2011.

## References

- Amgoud, L., and Besnard, P. (2009), 'Bridging the Gap Between Abstract Argumentation Systems and Logic', in *Proceedings of the 3rd International Conference on Scalable Uncertainty (SUM'09)* (Vol. 5785), Springer Lecture Notes in AI, Berlin: Springer Verlag, pp. 12–27.
- Amgoud, L., and Cayrol, C. (2002), 'A Model of Reasoning Based on the Production of Acceptable Arguments', *Annals of Mathematics and Artificial Intelligence*, 34, 197–215.
- Amgoud, L., and Vesic, S. (2011), 'A New Approach for Preference-Based Argumentation Frameworks', *Annals of Mathematics and Artificial Intelligence*, to appear.
- Baker, A., and Ginsberg, M. (1989), 'A Theorem Prover for Prioritized Circumscription', in *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 463–467.
- Baroni, P., and Giacomin, M. (2005), 'SCC-Recursiveness: A General Schema for Argumentation Semantics', *Artificial Intelligence*, 168, 162–210.
- Bench-Capon, T. (2003), 'Persuasion in Practical Argument Using Value-Based Argumentation Frameworks', *Journal of Logic and Computation*, 13, 429–448.
- Besnard, P., and Hunter, A. (2001), 'A Logic-Based Theory of Deductive Arguments', *Artificial Intelligence*, 128, 203–235.
- Besnard, P., and Hunter, A., (2008), *Elements of Argumentation*. Cambridge, MA: MIT Press.
- Bondarenko, A., Dung, P., Kowalski, R., and Toni, F. (1997), 'An Abstract, Argumentation-Theoretic Approach to Default Reasoning', *Artificial Intelligence*, 93, 63–101.
- Brewka, G. (1989), 'Preferred Subtheories: An Extended Logical Framework for Default Reasoning', in *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, San Mateo, CA: Morgan Kaufmann, pp. 1043–1048.
- Brewka, G., (1991), *Nonmonotonic Reasoning: Logical Foundations of Commonsense*, Cambridge: Cambridge University Press.
- Caminada, M. (2005), 'Contamination in Formal Argumentation Systems', in *Proceedings of the Seventeenth Belgian-Dutch Conference on Artificial Intelligence (BNAIC-05)*, Brussels, Belgium.

- Caminada, M. (2006), 'On the Issue of Reinstatement in Argumentation', in *Proceedings of the 11th European Conference on Logics in Artificial Intelligence (JELIA 2006)* (Vol. 4160), Springer Lecture Notes in AI, Berlin: Springer Verlag, pp. 111–123.
- Caminada, M., and Amgoud, L. (2007), 'On the Evaluation of Argumentation Formalisms', *Artificial Intelligence*, 171, 286–310.
- Chisholm, R., (1957), *Perceiving: A Philosophical Study*, Ithaca: Cornell University Press.
- Chisholm, R. (1974), 'Practical Reason and the Logic of Requirement', in *Practical Reason*, ed. S. Körner, Oxford: Blackwell Publishing Company, pp. 2–13.
- Doyle, J. (1979), 'Truth Maintenance Systems', *Artificial Intelligence*, 12, 231–272.
- Dung, P. (1995), 'On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and  $n$ -Person Games', *Artificial Intelligence*, 77, 321–357.
- Dung, P., Kowalski, R., and Toni, F. (2009), 'Assumption-Based Argumentation', in *Argumentation in Artificial Intelligence*, eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 199–218.
- Fahlman, S., (1979), *NETL: A System for Representing and Using Real-world Knowledge*, Cambridge, MA: MIT Press.
- Garcia, A., and Simari, G. (2004), 'Defeasible Logic Programming: An Argumentative Approach', *Theory and Practice of Logic Programming*, 4, 95–138.
- Ginsberg, M. (1994), 'AI and Nonmonotonic Reasoning', in *Handbook of Logic in Artificial Intelligence and Logic Programming*, eds. D. Gabbay, C. Hogger, and J. Robinson, Oxford: Clarendon Press, pp. 1–33.
- Gorogiannis, N., and Hunter, A. (2011), 'Instantiating Abstract Argumentation with Classical-Logic Arguments: Postulates and Properties', *Artificial Intelligence*, 175, 1479–1497.
- Hart, H. (1949), 'The Ascription of Responsibility and Rights', in *Proceedings of the Aristotelean Society*, Reprinted in *Logic and Language. First Series*, ed. A.G.N. Flew, 145–166. Oxford: Basil Blackwell, pp. 99–117.
- Horty, J., (2012), *Reasons as Defaults*, Oxford: Oxford University Press.
- Horty, J., and Thomason, R. (1988), 'Mixing Strict and Defeasible Inheritance', in *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-88)*, San Mateo, CA: Morgan Kaufmann, pp. 427–432.
- Horty, J., Thomason, R., and Touretzky, D. (1987), 'A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks', in *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, San Mateo, CA: Morgan Kaufmann, pp. 358–363.
- Horty, J., Thomason, R., and Touretzky, D. (1990), 'A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks', *Artificial Intelligence*, 42, 311–348.
- Israel, D. (1980), 'What's Wrong with Non-Monotonic Logic?', in *Proceedings of the First National Conference on Artificial Intelligence (AAAI-80)*, Stanford, CA: AAAI Press/MIT Press, pp. 99–101.
- Jakobovits, H. (2000), 'On the Theory of Argumentation Frameworks', Doctoral dissertation, Free University Brussels.
- Jakobovits, H., and Vermeir, D. (1999), 'Robust Semantics for Argumentation Frameworks', *Journal of Logic and Computation*, 9, 215–261.
- Krause, P., Ambler, S., Elvang-Gøransson, M., and Fox, J. (1995), 'A Logic of Argumentation for Reasoning Under Uncertainty', *Computational Intelligence*, 11, 113–131.
- Lin, F., and Shoham, Y. (1989), 'Argument Systems. A Uniform Basis for Nonmonotonic Reasoning', in *Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference*, San Mateo, CA: Morgan Kaufmann Publishers, pp. 245–255.
- Loui, R. (1987), 'Defeat Among Arguments: A System of Defeasible Inference', *Computational Intelligence*, 2, 100–106.
- Modgil, S. (2009), 'Reasoning About Preferences in Argumentation Frameworks', *Artificial Intelligence*, 173, 901–934.
- Modgil, S., and Prakken, H. (2011), 'Revisiting Preferences and Argumentation', in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, Menlo Park, CA: AAAI Press/IJCAI, pp. 1021–1026.
- Nute, D. (1988), 'Defeasible Reasoning: A Philosophical Analysis in Prolog', in *Aspects of Artificial Intelligence*, ed. J. Fetzer, Dordrecht: Kluwer Academic Publishers, pp. 251–288.

- Nute, D. (1994), 'Defeasible Logic', in *Handbook of Logic in Artificial Intelligence and Logic Programming*, eds. D. Gabbay, C. Hogger, and J. Robinson, Oxford: Clarendon Press, pp. 253–395.
- Parsons, S., Wooldridge, M., and Amgoud, L. (2003), 'Properties and Complexity of Some Formal Inter-Agent Dialogues', *Journal of Logic and Computation*, 13, 347–376.
- Pollock, J. (1970), 'The Structure of Epistemic Justification', in *Studies in the Theory of Knowledge*, American Philosophical Quarterly Monograph Series (Vol. 4), Oxford: Basil Blackwell Publisher, Inc., pp. 62–78.
- Pollock, J., (1974), *Knowledge and Justification*, Princeton: Princeton University Press.
- Pollock, J. (1987), 'Defeasible Reasoning', *Cognitive Science*, 11, 481–518.
- Pollock, J. (1992), 'How to Reason Defeasibly', *Artificial Intelligence*, 57, 1–42.
- Pollock, J. (1994), 'Justification and Defeat', *Artificial Intelligence*, 67, 377–408.
- Pollock, J., (1995), *Cognitive Carpentry. A Blueprint for How to Build a Person*, Cambridge, MA: MIT Press.
- Pollock, J. (1998), 'The Logical Foundations of Goal-Regression Planning in Autonomous Agents', *Artificial Intelligence*, 106, 267–335.
- Pollock, J. (1999), 'Planning Agents', in *Foundations of Rational Agency*, eds. M. Wooldridge and A. Rao, Dordrecht: Kluwer Academic Publishers.
- Pollock, J. (2002), 'Defeasible Reasoning with Variable Degrees of Justification', *Artificial Intelligence*, 133, 233–282.
- Pollock, J. (2005), 'Plans and Decisions', *Theory and Decision*, 57, 79–107.
- Pollock, J. (2007a), 'Reasoning and Probability', *Law, Probability and Risk*, 6, 43–58.
- Pollock, J. (2007b), 'Defeasible Reasoning', in *Reasoning: Studies of Human Inference and its Foundations*, eds. J. Adler and L. Rips, Cambridge University Press: Cambridge, pp. 451–470.
- Pollock, J. (2009), 'A Recursive Semantics for Defeasible Reasoning', in *Argumentation in Artificial Intelligence*, eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 173–197.
- Pollock, J. (2010), 'Defeasible Reasoning and Degrees of Justification', *Argument and Computation*, 1, 7–22.
- Poole, D. (1988), 'A Logical Framework for Default Reasoning', *Artificial Intelligence*, 36, 27–47.
- Prakken, H. (2010a), 'An Abstract Framework for Argumentation with Structured Arguments', *Argument and Computation*, 1, 93–124.
- Prakken, H. (2010b), 'On the Nature of Argument Schemes', in *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument*, eds. C. Reed and C. Tindale, London: College Publications, pp. 167–185.
- Prakken, H. (Forthcoming), 'Some Reflections on Two Current Trends in Formal Argumentation'.
- Prakken, H., and Sartor, G. (1997), 'Argument-Based Extended Logic Programming with Defeasible Priorities', *Journal of Applied Non-classical Logics*, 7, 25–75.
- Raz, J., (1975), *Practical Reason and Norms*, Princeton: Princeton University Press.
- Reiter, R. (1980), 'A Logic for Default Reasoning', *Artificial Intelligence*, 13, 81–132.
- Rescher, N., (1976), *Plausible Reasoning*, Assen: Van Gorcum.
- Rescher, N. (1977), *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*, Albany, NY: State University of New York Press.
- Ross, W. (1930), *The Right and the Good*, Oxford: Oxford University Press.
- Simari, G., and Loui, R. (1992), 'A Mathematical Treatment of Defeasible Argumentation and its Implementation', *Artificial Intelligence*, 53, 125–157.
- Touretzky, D. (1986), *The Mathematics of Inheritance Systems*, San Mateo, CA: Morgan Kaufmann.
- Touretzky, D., Horty, J., and Thomason, R. (1987), 'A Clash of Intuitions: The Current State of Nonmonotonic Multiple Inheritance Systems', in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, San Mateo, CA: Morgan Kaufmann, pp. 476–482.
- Vreeswijk, G. (1993), 'Studies in Defeasible Argumentation', Doctoral dissertation, Free University Amsterdam.
- Vreeswijk, G. (1997), 'Abstract Argumentation Systems', *Artificial Intelligence*, 90, 225–279.
- Walton, D. (1996), *Argumentation Schemes for Presumptive Reasoning*, Mahwah, NJ: Lawrence Erlbaum Associates.
- Walton, D., Reed, C., and Macagno, F., (2008), *Argumentation Schemes*, Cambridge: Cambridge University Press.