Arguments and Their Strength: Revisiting Pollock's Anti-Probabilistic Starting Points

Bart VERHEIJ
CodeX, Stanford University; Artificial Intelligence, University of Groningen

Abstract. Pollock's concepts of reasons and defeaters have been widely adopted, but his anti-probabilistic treatment of argument strength less so. After an explanation of how Pollock's concerns can be addressed probabilistically, the paper continues with a formal treatment of reasons, defeaters, and argument strength, while remaining within standard probability theory and its underlying classical logic. Pollock studied puzzles about self-defeat and collective defeat (associated with the lottery paradox), and it is shown how these can be addressed probabilistically. A normative framework for arguments and their strength, as provided here, is needed for the development of rationality support tools for the prevention of reasoning errors.

Keywords. Arguments, argument strength, probability, Pollock

1. Introduction

Pollock's approach [20,21] to the computational study of argument-based defeasible reasoning is a landmark in the field [2,24,28]. He distinguishes reasons attacking a conclusion from reasons attacking an inference step, referring to them as rebutting and undercutting defeaters, respectively. In his work, Pollock aimed to balance normative themes about which conclusions are justified given the available arguments, with computational concerns implied by the resource bounds of human and artificial reasoning. Dung [7] used Pollock's model as an illustration of the scope of his abstract model of argumentation, focusing on argument attack. Many of Pollock's main ideas have been incorporated in subsequent work. For instance, his rebutter-undercutter distinction is an ingredient of many argumentation models that extended the expressiveness of Dung's abstract model to include support [23,29,33,4].

There has been less attention for Pollock's positions on how arguments, logic and probability are related, although they are at the heart of his views. He studied and adapted ideas in classical logic, nonmonotonic logic and probability theory to fit the purposes of his OSCAR model. He strongly connected to nonmonotonic logic, noting that '[w]hat philosophers call defeasible reasoning is roughly the same as nonmonotonic reasoning in AI' [20, p. 481]. But he opposed what he called 'generic Bayesianism' [21, p. 95f.] and 'probabilism' [22, p. 11; published posthumously]. Generic Bayesianism is a theory in which 'reasons make their conclusions probable to varying degrees, and the ultimate conclusion is justified only if it is made sufficiently probable by the cumulative reasoning'. Probabilism is the view that degrees of justification obey the probability calculus. While he kept developing the specifics of his approach, his general views concerning the relations between arguments, logic and probability have remained quite stable between 1995 and 2010. In particular, he defended the following positions:[1]

1. Reasons can have different strengths, and conclusions can differ in their degree of justification. [21, p. 93][22, p. 8]
2. Degrees of justification do not work like probabilities. [21, p. 99][22, p. 11]

While the first of these two positions is hard to dispute, the second is more controversial, and depends strongly on Pollock's specific considerations. Two concerns stand out in his discussion. The first is discussed in his critique of generic Bayesianism [21, p. 95f.]: logically valid inference rules using multiple premises are not probabilistically valid. As an example, he mentions that P and Q logically entails $P \wedge Q$, while $\mathrm{p}(P \wedge Q)$ can be less than $\mathrm{p}(P)$ and than $\mathrm{p}(Q)$. In his interpretation, a Bayesian view therefore precludes inference from multiple premises, which is clearly needed in a model of reasoning.

Pollock's second concern is expressed in his discussion of probabilism [22, p. 11]: necessary truths have probability 1. Pollock claims that probabilism therefore implies that we would have been justified in believing Fermat's conjecture even before Andrew Wiles found a proof. For Pollock, this argument (that he refers to as simple and familiar) implies that probabilism is untenable.

Many of Pollock's specific positions about arguments, their strength and their structure depend on his anti-probabilistic stance. What would happen if we would embrace probability theory as a normative framework for his core notions? That is the question that inspired this paper. The goal of the paper is to revisit Pollock's anti-probabilistic starting points about arguments, their strength and their structure, and show that the key notions of reasons, defeaters, and argument strength can be formalized within the normative framework of probability theory. This goal implies that we will here not consider Pollock's extensive and insightful computational concerns, which have also determined many of his specific choices.

In the next section, Pollock's approach to argument-based defeasible reasoning is presented, in particular his notions of reasons and defeaters, and his views on argument strength. It is explained how Pollock's anti-probabilistic concerns can be addressed. In Section 3, we provide a formal treatment of reasons, defeaters, and argument strength, using probability theory and its underlying classical logic as a normative framework. Formal properties are investigated. In Section 4, we address two of Pollock's typical puzzles, on self-defeat and collective defeat. In Section 5, related research is discussed.

2. Elements of Pollock's approach to argument-based defeasible reasoning

Over the years Pollock has developed an elaborate approach to argument-based defeasible reasoning [20,21,22]. He kept extending the approach, and changed many details, including the arguments to support his choices. Here we focus on his concepts of reasons and defeaters (Section 2.1) and ideas about argument strength (Section 2.2). In Section 2.3, we explain how Pollock's anti-probabilistic concerns can be addressed.

[1] There are slight differences between Pollock's terminology in 1995, his terminology in 2010, and our paraphrasing. An attempt has been made to stay true to Pollock's ideas, while choosing terminology fitting the goals of this paper.

2.1. *Reasons and defeaters*

Pollock's theory of defeasible reasoning uses reasons as the building blocks of arguments. There are two kinds of reasons. Some reasons are *conclusive*. These are not defeasible and entail their conclusions. Other reasons are *prima facie*. They create a presumption for their conclusion and may be defeated. Using a classical propositional logical language, reasons can be written as pairs (φ, ψ), where φ expresses the premises of the reason, conjunctively combined, and ψ the conclusion.[2]

Defeaters are a special kind of reasons, namely reasons that defeat prima facie reasons. Pollock dinstinguishes two kinds of defeaters. The first kind are *rebutting* defeaters that deny the conclusion of a prima facie reason. So when (φ, ψ) is a prima facie reason, a reason $(\varphi', \neg\psi)$ for the negated conclusion is a rebutting defeater. The second kind are *undercutting* defeaters that attack the connection between premises and conclusion of a reason. Pollock models undercutters by nesting the negation of a kind of conditional in the conclusion of a reason: When (φ, ψ) is a prima facie reason, a reason $(\varphi', \neg(\varphi, \psi))$ expresses an undercutting defeater.[3]

2.2. *Argument strength*

Pollock introduces strength in his model when discussing his statistical syllogism:

> If $r > 0.5$, then $\varphi \wedge \mathrm{p}(\psi|\varphi) \geq r$ is a prima facie reason for ψ. [21, p. 93][22, p. 9]

Here $\mathrm{p}(\psi|\varphi)$ denotes the conditional probability of ψ given φ, which can be expressed in Pollock's logical language. According to Pollock, the strength of a reason based on the statistical syllogism should be a monotonic increasing function of r. In [21], he mentions the possibility of using the lower bound r as the strength of such a reason, but finalizes on a different proposal (p. 94). In [22], no explicit measure of strength for the reasons based on the statistical syllogism is given.

In Pollock's approach, reasons are combined in larger structures, referred to as arguments and inference graphs, and he considers how the degree of justification of a conclusion depends on these. He addresses the role of strengths of individual reasons using the Weakest Link Principle. For deductive arguments, this principle reads that 'The degree of support of the conclusion of a deductive argument is the minimum of the degrees of support of its premises'.[21, p. 99], and formalized for conjunctions (cf. [22, p. 17]):

$$\text{degree-of-justification}(\varphi \wedge \psi) = \min\{\text{degree-of-justification}(\varphi), \text{degree-of-justification}(\psi)\}$$

A weaker constraint would be that the minimum of the degrees of justification of the conjuncts is an upper bound for that of the conjunction, a constraint with a valid counterpart in probability theory. Pollock strengthens that constraint beyond what is probabilistically valid in order to be able to quickly compute the degree of justification of conjunctions.

Pollock has kept developing his proposals of mechanisms that determine how the constellations of reasons and their strengths determine which conclusions are justified. These mechanisms, and the almost experimental style of trying and testing them, are important and influential, but their specifics are not needed for the rest of this paper.

---

[2] Pollock models reasons as pairs (Γ, p), where Γ is the set of premises of the reason, and p the conclusion.

[3] Pollock writes $(\Gamma', \neg\lceil\Pi\Gamma >> p\rceil)$, where (Γ, p) is a prima facie reason, $\Pi\Gamma$ denotes the conjunction of the premises, and $\lceil\Pi\Gamma >> p\rceil$ the connection between premises and conclusion.

2.3. *Addressing Pollock's concerns*

Pollock's first anti-probabilistic concern is that logically valid inference rules using multiple premises are not probabilistically valid. The second is that necessary truths have probability 1. How can these concerns be addressed, within a probabilistic setting?

The first concern is addressed by accepting that the conjunction of presumptive conclusions is not in general a valid step to make. Only when making conclusive reasoning steps within a coherent position, conclusions can be safely combined by conjunction. As soon as reasoning becomes presumptive, it is possible that different conclusions are incompatible with each other, and cannot be conjunctively combined. For instance, when a witness implicates a suspect of a crime (φ) and a second witness another suspect (ψ), each of φ and ψ is presumptively supported, but the conjunction $\varphi \wedge \psi$ is not. Formally, this means that what is called the (And)-rule should not be accepted (see Section 3.3).

Pollock's second concern is addressed by focusing on the normative setting of arguments and their strength. Normatively, the fact that necessary truths have probability 1 is unproblematic. By distinguishing actual justification from normative justification, we see that, normatively, Fermat's conjecture would have been justified during the centuries before Andrew Wiles' proof, but the actual justification only came about when Wiles discovered a proof. In terms of argument strength, the proof of a necessary truth is an argument of strength 1, that leaves no uncertainty.

3. Probability theory as a normative framework for arguments and their strength

In Pollock's approach, arguments are built from reasons and defeaters. In this section, we formalize reasons, defeaters, and argument strength, using standard probability theory as normative framework. Formal properties are studied.

3.1. *Standard probability and its underlying classical logic*

We remain within standard probability theory and its underlying classical logic. We use a logical language L with binary connectives $\wedge$ and $\vee$, expressing conjunction and disjunction, respectively, and a unary connective $\neg$, expressing (classical) negation. The language is constructed inductively from a non-empty set of propositional Boolean variables P, with an associated classical deductive consequence relation $\models$. Sentences φ and $\psi \in L$ are said to be logically incompatible when $\models \neg(\varphi \wedge \psi)$. A sentence $\varphi \in L$ is a logical truth when $\models \varphi$.

Probability functions are real-valued functions governed by the standard Kolmogorov axioms [12]:

1. $\mathrm{p}(\varphi) \geq 0$ for all $\varphi \in L$.
2. If $\varphi \in L$ is a logical truth, then $\mathrm{p}(\varphi) = 1$.
3. $\mathrm{p}(\varphi \vee \psi) = \mathrm{p}(\varphi) + \mathrm{p}(\psi)$ for all φ and $\psi \in L$ such that φ and ψ are logically incompatible.[4]

Conditional probabilities play a central role in the following: For sentences φ and $\psi \in L$, with $\mathrm{p}(\varphi) > 0$, the conditional probability $\mathrm{p}(\psi|\varphi)$ is defined as $\mathrm{p}(\varphi \wedge \psi)/\mathrm{p}(\varphi)$.

[4] Countably infinite additivity is left out of the discussions in this paper.

3.2. Reasons and defeaters

A reason is a pair (φ, ψ), with φ and ψ sentences in the language L. The premises of the reason are expressed by φ, and the conclusions by ψ. The plural is used — for premises and conclusions — as they typically consist of more than one element combined by conjunction. For a reason (φ, ψ), we will refer to the conjunction of premises and conclusions $\varphi \wedge \psi$ as the *case made* by the reason (a notion introduced in [30,31]).

Whether a reason is conclusive or prima facie, depends on a probability function that serves as a model. We assume a probability function p, and define:

> A reason (φ, ψ) is *conclusive* if and only if $\mathrm{p}(\psi|\varphi) = 1$.
> A reason (φ, ψ) is *prima facie* if and only if $\mathrm{p}(\psi|\varphi) > 0$.
> φ' is the *relevant reason* for ψ in (φ, ψ) if and only if φ' is logically maximally general such that (φ', φ) is conclusive.

The probability of 1 of a conclusive reason (φ, ψ) expresses that all φ-events are ψ-events. The positive probability of a prima facie reason (φ, ψ) expresses that a positive proportion of φ-events are also ψ-events. A prima facie reason indicates the possibility of the case $\varphi \wedge \psi$ made by the reason.[5] Relevant reasons single out that part of the premises that determine the strength of the reason for its conclusions.

A prima facie reason can be so strong that it is actually conclusive. By the probability calculus, a prima facie reason (φ, ψ) is conclusive if and only if $(\varphi, \neg\psi)$ is not a prima facie reason. So an inconclusive prima facie reason (φ, ψ) has an associated prima facie reason $(\varphi, \neg\psi)$ for the opposite conclusion. When (φ, ψ) is a weak prima facie reason, say of strength ε close to zero, $(\varphi, \neg\psi)$ is strong, with a strength $1 - \varepsilon$, close to 1. One may think that this implies that a weak alibi provides a strong reason for guilt, which would be strange. However, a weak alibi provides a weak reason for innocence on the basis of the alibi, hence a strong reason against innocence on the basis of the alibi. There can be alternative scenarios, some implying innocence, others guilt.

The strength of a prima facie reason (φ, ψ) can change when adding premises to include circumstances χ. The result can be defeat:

> A prima facie reason (φ, ψ) is *defeated in the circumstances* χ if and only if $(\varphi \wedge \chi, \psi)$ is not a prima facie reason. The reason $(\varphi \wedge \chi, \psi)$ is the *defeater*.

It follows that there are two kinds of defeaters. In the first kind, the premises $\varphi \wedge \chi$ of the defeater have probability 0, and the defeater has no well-defined strength. The reason is defeated by falsifying its premises. In the second kind, the defeater's premises have positive probability, and the defeater has strength 0. Then the prima facie reason (φ, ψ) is defeated by circumstances χ for which $\mathrm{p}(\varphi \wedge \psi \wedge \chi) = 0$. In other words, the circumstances χ make the case $\varphi \wedge \psi$ made by the reason impossible. Such defeaters break the connection between the premises and conclusions of the defeated reason; cf. Pollock's undercutting defeaters. Also, by the probability calculus, such a defeater $(\varphi \wedge \chi, \psi)$ corresponds to a conclusive reason $(\varphi \wedge \chi, \neg\psi)$ for the opposite conclusion; cf. Pollock's rebutting defeaters. Hence, in the terminology proposed here, given a probability function, defeaters of the second kind are both undercutting and rebutting. This is a consequence of the combination of a liberal notion of prima facie reasons with a strict notion of defeaters.

[5] Our use of the term possibility as positive probability is different from its use in possibility theory [6].

3.3. *Properties of prima facie reasons*

By our characterization of prima facie reasons, they come in different strengths, as measured by the conditional probability of the conclusions given the premises. Prima facie reasons can be strong with a strength close to or equal to 1, and prima facie reasons can be weak with a strength close to 0. We write $s(\varphi, \psi)$ for the strength $\mathrm{p}(\psi|\varphi)$ of the reason (φ, ψ). Using this notation, $s(\top, \varphi) = \mathrm{p}(\varphi)$, and the definition of conditional probability can be rewritten as $s(\varphi, \psi) = s(\top, \varphi \wedge \psi)/s(\top, \varphi)$, where $\top$ denotes a logical necessity.

The strengths of prima facie reasons are governed by Bayes' rule. We have the following rephrasing of Bayes' rule:

$$\frac{s(\varphi, \psi)}{s(\top, \psi)} = \frac{s(\psi, \varphi)}{s(\top, \varphi)}$$

The well-known likelihood ratio formula connects the prior odds of two hypothetical positions with the posterior odds in light of new information; in the notation of strengths:

$$\frac{s(\varphi, \psi)}{s(\varphi, \chi)} = \frac{s(\psi, \varphi)}{s(\chi, \varphi)} \cdot \frac{s(\top, \psi)}{s(\top, \chi)}$$

Bayes' rule and the likelihood ratio formula govern how the strengths of a reason and its 'reverse' — arrived at by switching premises and conclusions — are connected. Switching to the reverse of a conditional probability is a common and well-known error. Ignoring strength, however, reversal is not erroneous:

Proposition 3.1. *Let p be a probability function. Then, for all sentences φ and $\psi \in L$: (φ, ψ) is a prima facie reason if and only if (ψ, φ) is a prima facie reason.*

The strengths of (φ, ψ) and (ψ, φ) can be very different, and are as said connected by Bayes' rule. A more general property ties a prima facie reason to the case made by it:

Proposition 3.2. *Let p be a probability function. Then, for all sentences φ and $\psi \in L$: (φ, ψ) is a prima facie reason if and only if $(\top, \varphi \wedge \psi)$ is a prima facie reason.*

The proposition makes explicit that the existence of a prima facie reason corresponds to the possibility of the case made by the reason, expressed by a positive probability of the case. Typically, the 'a priori' support for the case made will be (much) lower than that of the conclusion by the reason. Formally, we have that $s(\top, \varphi \wedge \psi) \leq s(\varphi, \psi)$.

The next property is a slight, but useful, generalisation:

Proposition 3.3. *If (φ, ψ) is a prima facie reason, and α and β are logical consequences of the case $\varphi \wedge \psi$ made by the reason, then (α, β) is also a prima facie reason.*

The property shows that a prima facie reason implies the existence of many other prima facie reasons; and quite liberally so. For instance, the property implies that when (φ, ψ) is a prima facie reason, $(\top, \psi)$ is too. In a sense, the premises of a prima facie reason can be ignored, typically at the price of a significant reduction of the strength of the reason.

We list some properties, as studied in the theory of nonmonotonic consequence relations [19], here focusing on those in [30,31]:

Proposition 3.4. 1. *If* (φ, ψ) *is a prima facie reason,* $\models \varphi \leftrightarrow \varphi'$ *and* $\models \psi \leftrightarrow \psi'$*, then* (φ', ψ') *is a prima facie reason. (Logical Equivalence)*
2. *If* (φ, ψ) *is a prima facie reason, then* $(\varphi, \varphi \wedge \psi)$ *is a prima facie reason. (Antecedence)*
3. *If* $(\varphi, \varphi \wedge \psi)$ *is a prima facie reason, then* (φ, ψ) *is a prima facie reason. (Premise Reduction)*
4. *If* $(\varphi, \psi \wedge \chi)$ *is a prima facie reason, then* (φ, ψ) *is a prima facie reason. (Right Weakening)*
5. *If* $(\varphi, \psi \wedge \chi)$ *is a prima facie reason, then* $(\varphi \wedge \psi, \chi)$ *is a prima facie reason. (Conjunctive Cautious Monotony)*
6. *If* (φ, ψ) *and* $(\varphi \wedge \psi, \chi)$ *are prima facie reasons, then* $(\varphi, \psi \wedge \chi)$ *is a prima facie reason. (Conjunctive Cumulative Transitivity)*

Still, some commonly considered properties do not hold. (Reflexivity) does not hold: (φ, φ) is only a prima facie reason when φ is possible, i.e., has positive probability. For instance, logical inconsistencies, denoted $\perp$, are not possible, so $(\perp, \perp)$ is not a prima facie reason. More generally, $(\varphi, \perp)$ cannot be a prima facie reason, since $\mathrm{p}(\perp) = 0$, giving strength 0. The reason $(\perp, \varphi)$ cannot be prima facie since it has undefined strength.

As discussed (Section 2.3), the (And)-property should not hold, and indeed it does not. When (φ, ψ) and (φ, χ) are prima facie reasons, it need not be the case that $(\varphi, \psi \wedge \chi)$ is a prima facie reason. Conflicting reasons are an example: $(\varphi, \psi \wedge \neg\psi)$ is never a prima facie reason, not even when (φ, ψ) and $(\varphi, \neg\psi)$ are. The (Or)-property does hold: when (φ, ψ) and (φ', ψ) are prima facie reasons, then $(\varphi \vee \varphi', \psi)$ is.

The next proposition connects prima facie and conclusive reasons (part 1), and shows a property close to defeasible Modus ponens (part 2): reasons can be chained, unless the case supported by the first step provides a defeater for the second step.

Proposition 3.5. 1. *If* (φ, ψ) *is a prima facie reason, and* $(\varphi, \neg\psi)$ *is not, then* (φ, ψ) *is a conclusive reason.*
2. *If* (φ, ψ) *and* (ψ', χ) *are prima facie reasons with* $\varphi \wedge \psi \models \psi'$*, then* $(\varphi, \psi \wedge \chi)$ *is a prima facie reason, unless* $(\varphi \wedge \psi, \chi)$ *is a defeater for* (ψ', χ)*.*

In a critical discussion, a prima facie reason (φ, ψ), covering one's overall starting points φ and positions ψ, can be developed in four natural ways:

1. *Strengthening the premises*: from (φ, ψ) to $(\varphi \wedge \chi, \psi)$.
This occurs when new evidence comes in, or when, more generally, more specific information is used as starting point of the reasoning. When strengthening the premises, the strength of a prima facie reason can increase (up to becoming 1), decrease (up to becoming 0), or remain the same. If (φ, ψ) is a prima facie reason, $(\varphi \wedge \chi, \psi)$ may not be. If (φ, ψ) is a conclusive reason, $(\varphi \wedge \chi, \psi)$ will also be, as long as the premises $\varphi \wedge \chi$ are possible.
2. *Strengthening the conclusions*: from (φ, ψ) to $(\varphi, \psi \wedge \chi)$.
This occurs when a rule is applied, or when, more generally, more specific conclusions are drawn. When strengthening the conclusions, the strength of a prima facie reason cannot increase, since probability theory implies $s(\varphi, \psi) \geq s(\varphi, \psi \wedge \chi)$. The strength can decrease (up to becoming 0), or remain the same. If (φ, ψ) is a prima facie reason, $(\varphi, \psi \wedge \chi)$ may not be. If (φ, ψ) is a conclusive reason, $(\varphi, \psi \wedge \chi)$ may not be, and may not even be a prima facie reason.

3. *Weakening the premises*: from (φ, ψ) to $(\varphi \vee \chi, \psi)$.
   This occurs when the premises are questioned, or when, more generally, the starting points of reasoning are generalized. When weakening the premises, the strength of a prima facie reason can increase, but not from a value below 1 to equal to 1: if (φ, ψ) is not conclusive, we find that $\mathrm{p}(\neg\psi|\varphi) > 0$, hence $\mathrm{p}(\neg\psi|\varphi \vee \psi) > 0$. The strength can decrease (but not become 0), and can remain the same. If (φ, ψ) is a prima facie reason, $(\varphi \vee \chi, \psi)$ is too. If (φ, ψ) is a conclusive reason, $(\varphi \vee \chi, \psi)$ may not be, but it will be a prima facie reason.
4. *Weakening the conclusions*: from (φ, ψ) to $(\varphi, \psi \vee \chi)$.
   This occurs when conclusions are retracted to a safer position, or when, more generally, the conclusions are generalized. When weakening the conclusions, the strength of a prima facie reason can increase (up to becoming 1), but cannot decrease since probability theory implies that $s(\varphi, \psi) \leq s(\varphi, \psi \vee \chi)$. The strength can remain the same. If (φ, ψ) is a prima facie reason, $(\varphi, \psi \vee \chi)$ is too. If (φ, ψ) is a conclusive reason, $(\varphi, \psi \vee \chi)$ is too.

Typically, the starting points and positions in a critical discussion leave room for uncertainty. This corresponds to (φ, ψ) having strength below 1. Developing the starting points and positions may remove all uncertainty, i.e., lead to (φ, ψ) with strength 1. When a weaker burden of proof than certainty is in place, a strength below 1 considered sufficiently strong can also be the successful outcome of a critial discussion.

3.4. Comparison to Pollock's approach to arguments and their strength

Our idea was to embrace standard probability theory, and formalize the notions of reasons, defeaters, and argument strength within a probabilistic setting. Since we use probability functions as models that provide reasons (and defeaters), the reasons come with a natural measure of strength, the conditional probability of the conclusions given the reasons. In Pollock's approach, strength is associated with the statistical syllogism that leads him to consider reasons that involve statements about probabilities, such as $\varphi \wedge \mathrm{p}(\psi|\varphi) \geq r$ as a reason for ψ. One role of such meta-reasons in Pollock's approach is that it allows for the separatation of argument strength from the conditional probabilities related to them. This was needed because of his anti-probabilistic views. In the probabilistic perspective here, such separation is not needed. Only reasons expressed in the logical language L are considered, and not also statements about probabilities.

We saw that Pollock's statistical syllogism selected sufficiently strong reasons. Only when $r > 0.5$ with r the lowerbound of the conditional probability, the syllogism gives rise to a prima facie reason $\varphi \wedge \mathrm{p}(\psi|\varphi) \geq r$ for ψ. In his setting, focusing on resource-bounded computation, it is natural to restrict to sufficiently strong reasons. The normative approach presented here includes weaker reasons, even reasons with a strength close to 0. An advantage is that all options, even weakly supported ones, are a part of the reasoning. This is useful, even normatively necessary since, when new information comes in, weakly supported options can become strongly supported, up to the point of certainty (as we saw in the discussion of strengthening the premises in Section 3.3).

Pollock used a weakest link principle to combine the degrees of justification of reasons for separate conclusions. If premises φ provide a prima facie reason for ψ and for χ, the analog of Pollock's weakest link principle would be that the strength of $(\varphi, \psi \wedge \chi)$ is equal to the minimum of the strengths of (φ, ψ) and (φ, χ). That principle is not

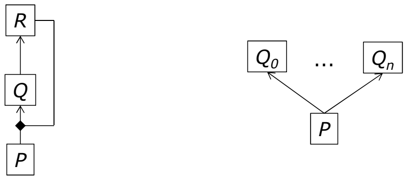

Figure 1. Two of Pollock's puzzles: self-defeat and collective defeat

endorsed in the present proposal, as its probabilistic counterpart is invalid. As we saw (Section 3.3), the strength of $(\varphi, \psi \wedge \chi)$ can be any value equal to or lower than (φ, ψ), including 0. Consider, for instance, a situation in which φ is a prima facie reason for ψ of strength 0.7. Then φ is a prima facie reason for $\neg\psi$ of strength 0.3. The weakest link principle would then give strength 0.3 to the reason $(\varphi, \psi \wedge \neg\psi)$. Probability theory attaches the strength 0 to this reason, implying that $(\varphi, \psi \wedge \neg\psi)$ is not even a prima facie reason. In [22], Pollock considers situations in which the strength of reasons is diminished, referring to the associated reasons as diminishers. He mentions that these require a principle other than the weakest link principle. Normatively, probability theory provides a sensible handling of diminishing: χ is a diminisher for a prima facie reason (φ, ψ) if the strength of (φ, ψ) is higher than that of $(\varphi \wedge \chi, \psi)$.

In Pollock's approach, reasons (and defeaters) are used to construct arguments or inference graphs, and he then develops criteria that determine the degree of justification of the conclusions that follow on the basis of the structure of the graph. Proposition 3.3 shows how, in the present approach, reasons contain subreasons. Proposition 3.5 shows how reasons can be chained.

4. Two puzzles

Several puzzles play a prominent role in Pollock's discussions of reasons and the arguments constructed from them. We discuss two, and show how they are addressed in the probabilistic approach here. The first concerns self-defeat, and uses only qualitative reasoning, since it does not involve the strength of the reasons, except for the extreme values 0 and 1. The second concerns collective defeat, and uses quantitative reasoning, since it has a role for the strength of reasons in between 0 and 1.

The self-defeat puzzle[6] involves a prima facie reason P for Q which is in turn a prima facie reason for R. R is an undercutting defeater attacking P as a reason for Q (Figure 1, on the left). Given P, is R justified or not? Assume R is justified, then that should be on the basis of Q, but if R is justified there is no reason justifying Q since R undercuts the only possible reason P for Q; contradiction. Assume now that R is not justified. But then Q cannot be a reason for R. But that is only the case when P as a reason for Q is undercut, requiring that R is justified; contradiction. Pollock uses this paradoxical outcome to test and adapt his proposals to determine the justification status of conclusions.

Modeling this example probabilistically as in this paper, we get the following. The undercutting defeater R attacking P as a reason for Q implies that the reason $(P \wedge R, Q)$

[6]See [21, p. 118–119]. By the specifics of Pollock's and our rephrasing, his Figure 9 looks different from the figure used here.

is not a prima facie reason, i.e., $p(Q|P \wedge R) = 0$. But then by the probability calculus also $p(R|P \wedge Q) = 0$, so P is an undercutting defeater attacking Q as a reason for R. Also by the probability calculus, $p(Q \wedge R|P) = 0$. As a result, the paradox disappears. P is a prima facie reason for Q, but not for $Q \wedge R$, so the step from P via Q to R cannot be made.

The second puzzle concerns collective defeat and is associated with the lottery paradox [21, p. 112]. There is a reason P for $n+1$ different, incompatible conclusions $Q_0, \ldots, Q_n$ (Figure 1, on the right). For instance, consider a fair lottery with a million participants and one winner, described in the premises P, and let Q_i express that participant i will *not* win. Then P provides a strong but inconclusive prima facie reason for each of the Q_i. The puzzle is that this suggests that P is a prima facie reason for the conjunction of the Q_i. However then there is a contradiction with P, since someone will win. Pollock proposes a principle of collective defeat, that ensures that each of the Q_i is defeated.[7]

In the setting presented here, the puzzle disappears, as the (And) rule that allows for the conjunction of different conclusions is not endorsed. Each of the Q_i is strongly supported by P, but the strength of the reason from P to their conjunction, a contradiction, is 0. In this analysis of the puzzle, there can be several positions that are each well-supported, while their conjunction is not.

5. Discussion of related work

Contemporary formal and computational models of argument typically use Dung's abstract semantics [7] as normative framework. Dung investigated four semantics (grounded, complete, preferred, stable), after which several variations have been proposed [1]. Extending expressiveness to also include argument support leads to further variations (e.g., [29, p. 341] lists eleven variations, some more reasonable than others). The wealth of options has inspired work on the normative principles governing the different semantics [5], or has been accepted as a model parameter [23]. Typically, argument strength, which is at the heart of Pollock's concerns, is not investigated in abstract argumentation research. Recently, Hunter [15], inspired by [17,8,26], has studied abstract argumentation probabilistically, connecting to the model of deductive arguments he proposed with Besnard [3]. In this work, Pollock's specific concerns are not addressed. Formal bridges between abstract argumentation and probabilities are built, but probability theory is not used as the guiding normative framework as in this paper. Also the connections between argumentation and Bayesian Networks have become the subject of study, proposing idioms for the design of networks that incorporate arguments and hypothetical scenarios [14,9,32] and studying how to extract support and attack relations [27].

The normative underpinning of Pollock's computational model of arguments and their strength has been addressed before, in particular in epistemology. For instance, Fitelson [10] discusses elements of Pollock's work focusing on philosophical considerations in Bayesian epistemology. He does not reconstruct Pollock's central notions of reasons and defeaters, nor discusses puzzles such as self-defeat and collective defeat, so typical for Pollock. Spohn [25] contrasts Pollock's approach with his theory of ranking functions, a third alternative in formal epistemology, next to probability. Another formal-philosophical study of Bayesian epistemology is [18], where the lottery paradox is discussed, in connection with Kyburg, but not with Pollock.

[7] Kyburg introduced the lottery paradox in the 1960s; see, e.g., [16]. It led to extensive scholarly debate.

[11] pioneered the formal connections between argumentation, nonmonotonic reasoning and probabilities. More recently, [13] continued the study of the general properties of nonmonotonic reasoning in relation with probability theory. Retaking an argumentation perspective, [31] does not endorse the (And)-rule (as this paper), and proposes a theory that is compatible with probability theory, but also with non-standard theories. In contrast, this paper interprets arguments and their strength within standard probability theory.

6. Conclusion

By the design of his OSCAR model, Pollock aimed to address normative and computational concerns. His investigation of reason-based, ampliative, defeasible reasoning led him to deviate from classical logic and probability theory.

In the present paper, we have shown that the notions of reasons and defeaters can be interpreted while remaining within standard probability theory with its underlying classical logic. Strength is formally modeled as the conditional probability of the conclusions given the premises. Formal properties of prima facie reasons have been studied, and it has been explained how Pollock's anti-probabilistic concerns are addressed. The proposed formalization of prima facie reasons is liberal, and includes (very) weak reasons. Normatively, this is necessary since weak reasons can become strong, even conclusive, when adding premises. The counterpart notion of defeaters is strict.

The presented approach combines qualitative reasoning in terms of reasons and defeaters, with quantitative reasoning using argument strength. This is required since, for instance, in forensic evidential reasoning, the qualitative reasoning typical for judges and juries (who typically use arguments and scenarios) must be combined with the quantitative reasoning of forensic experts (in terms of statistics).[8] The normative perspective on arguments and their strength, proposed in this paper, can serve as the formal foundation of rationality support tools that help prevent reasoning errors.

References

[1] P. Baroni, M. Caminada, and M. Giacomin. Review: an introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.

[2] T.J.M. Bench-Capon and P.E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[3] P. Besnard and A. Hunter. Argumentation based on classical logic. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 133–152. Springer, Berlin, 2009.

[4] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[5] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[6] D. Dubois and H. Prade. Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32(1):35–66, 2001.

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

[8] Cf. the NWO Forensic Science project on arguments, scenarios and Bayesian Networks; `http://www.ai.rug.nl/~verheij/nwofs/`.

[8] P.M. Dung and P. Thang. Towards probabilistic argumentation for jury-based dispute resolution. In *Computational Models of Argument: Proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8-10, 2010*, pages 171–182. IOS Press, Amsterdam, 2010.

[9] N. Fenton, M. Neil, and D.A. Lagnado. A general structure for legal arguments about evidence using Bayesian Networks. *Cognitive Science*, 37:61–102, 2013.

[10] B. Fitelson. Pollock on probability in epistemology. *Philosophical Studies*, 148:455–465, 2010.

[11] H. Geffner and J. Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, 53(2-3):209–244, 1992.

[12] A. Hájek. Interpretations of probability. In E.N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, 2011.

[13] J. Hawthorne and D. Makinson. The quantitative/qualitative watershed for rules of uncertain inference. *Studia Logica*, 86(2):247–297, 2007.

[14] A.B. Hepler, A.P. Dawid, and V. Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6(1-4):275–293, 2007.

[15] A. Hunter. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55:607–638, 2014.

[16] H. Kyburg. The rule of adjunction and reasonable inference. *Journal of Philosophy*, 94:109–125, 1997.

[17] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Proceedings of the First International Workshop on the Theory and Applications of Formal Argumentation (TAFA 2011)*. Springer, Berlin, 2011.

[18] H. Lin and K.T. Kelly. Propositional reasoning that tracks probabilistic reasoning. *Journal of Philosophical Logic*, 41:957–981, 2012.

[19] D. Makinson. General patterns in nonmonotonic reasoning. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Nonmonotonic Reasoning and Uncertain Reasoning*, pages 35–110. Clarendon Press, Oxford, 1994.

[20] J.L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.

[21] J.L. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. The MIT Press, Cambridge (Massachusetts), 1995.

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

[23] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[24] I. Rahwan and G.R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, Dordrecht, 2009.

[25] W. Spohn. A brief comparison of pollock's defeasible reasoning and ranking functions. *Synthese*, 131(1):39–56, 2002.

[26] M. Thimm. A probabilistic semantics for abstract argumentation. In *Proceedings of the European Conference on Artificial Intelligence (ECAI 2012)*, pages 750–755. IOS Press, Amsterdam, 2012.

[27] S. Timmer, J.J. Meyer, H. Prakken, S. Renooij, and B. Verheij. Inference and attack in Bayesian Networks. In K. Hindriks, M. de Weerdt, B. van Riemsdijk, and M. Warnier, editors, *25th Benelux Conference on Artificial Intelligence (BNAIC 2013)*, pages 199–206. Delft University, 2013.

[28] F.H. van Eemeren, B. Garssen, E.C.W. Krabbe, A.F. Snoeck Henkemans, B. Verheij, and J.H.M. Wagemans. Chapter 11: Argumentation in Artificial Intelligence. In *Handbook of Argumentation Theory. A Comprehensive Overview of the State of the Art*. Springer, Berlin, to appear 2014.

[29] B. Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346, 2003.

[30] B. Verheij. Argumentation and rules with exceptions. In B. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument: Proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8-10, 2010*, pages 455–462. IOS Press, Amsterdam, 2010.

[31] B. Verheij. Jumping to conclusions. a logico-probabilistic foundation for defeasible rule-based arguments. In L. Fariñas del Cerro, A. Herzig, and J. Mengin, editors, *13th European Conference on Logics in Artificial Intelligence, JELIA 2012. Toulouse, France, September 2012. Proceedings (LNAI 7519)*, pages 411–423. Springer, Berlin, 2012.

[32] C. Vlek, H. Prakken, S. Renooij, and B. Verheij. Modeling crime scenarios in a Bayesian Network. In *The 14th International Conference on Artificial Intelligence and Law (ICAIL 2013). Proceedings of the Conference*, pages 150–159. ACM Press, New York (New York), 2013.

[33] G. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.