

Bayesian Argument Strength and Logics of Argumentation

Henry Prakken

*Department of Information and Computing Sciences, Faculty of Science,
Utrecht University, The Netherlands.*

9 March 2026

Abstract Recently, Bayesian approaches to argument strength have become popular in argumentation theory and psychology. A parallel development in artificial intelligence is the incorporation of probabilities in formal models of argumentation. Among other things, it has been studied how arguments can be given probabilistic strength in the well-known ASPIC+ framework for structured argumentation. This paper discusses the relation between these two developments, in particular how the probabilistic version of ASPIC+ can further clarify and extend the Bayesian approaches to argument strength in argumentation theory and psychology.

1 Introduction

In the last 20 years a Bayesian approach to argument strength has become popular in argumentation theory and psychology (Hahn & Oaksford 2006, Zenker 2012, Hahn & Hornikx 2016). The core idea is to define the strength of an argument as the conditional probability of its conclusion given its premises. Then related notions can also be defined in probabilistic terms. For example, an argument's force can be defined as the degree to which its premises increases the probability of its conclusion, in other words, as the difference between the argument's strength and the unconditional probability of its conclusion. One way to determine conditional probabilities is with Bayes' theorem, and this is why this approach to argument strength is often called 'Bayesian'.

A parallel development in artificial intelligence is the incorporation of probabilities in formal models, or logics, of argumentation (Hunter et al. 2021). This has mainly been done with Dung's (1995) theory of so-called abstract argumentation frameworks, which frameworks are nothing else but a given set of arguments with a binary notion of defeat¹, with no assumption whatsoever made on either how arguments look like or when and how they defeat each other. While Dung's frameworks have proved useful in unifying several aspects of defeasible reasoning and argumentation, they are by themselves too abstract to model many interesting aspects of argumentation, including probabilistic aspects. For this reason, I have in Prakken (2018b) studied how arguments can be given probabilistic strength in the ASPIC+ framework (Prakken 2010, Modgil & Prakken 2014, 2018), which gives structure to the arguments in Dung's frameworks and which defines when and how they defeat each other.

¹ Dung used the term 'attack' for this relation but for reasons explained below I rename it to defeat.

The aim of the present paper is twofold. The first is to make the ideas in my 2018b paper accessible for scholars in argumentation theory, and the second is to relate these ideas to the research on Bayesian argument strength in argumentation theory and psychology.

This paper is organised as follows. In Section 2 I summarise the main formalisms assumed in my paper: the basics of probability theory, the ASPIC+ framework and Dung’s theory of abstract argumentation frameworks. In Section 3 I sketch how I see the relation between three kinds of argument strength: logical, dialectical and rhetorical argument strength. This sketch serves as the background of Section 4, in which I present the core ideas of my 2018b paper on modelling probabilistic argument strength in ASPIC+ and discuss how it relates to Bayesian accounts of argument strength in argumentation theory and psychology. I conclude in Section 5.

2 Preliminaries

In this section I review the basics of probability theory, the ASPIC+ framework and Dung’s theory of abstract argumentation frameworks

2.1 Probability theory²

In probability theory, uncertainty concerning the truth of a statement is expressed in a number between 0 and 1 (or equivalently between 0% and 100%). A probability of 1 (or 100%) means that the statement is certainly true, a probability of 0 (or 0%) means that it is certainly false, and every number or percentage between these extremes expresses a degree of uncertainty. An important property of probabilities is that if two statements cannot be true at the same time and together exhaust all possibilities, then they add up to 1 (or to 100%). For example, if the probability that it will rain tomorrow is 0.8 (or 80%) then the probability that it will not rain tomorrow is 0.2 (or 20%).

Probabilities can also be conditional. A conditional probability expresses the probability that some statement is true given that, or assuming that some other statement is true. Conditional probabilities have much the same properties as unconditional probabilities. They are also between 0 and 1 (or between 0% and 100%), and when two statements A and B exclude each other and are jointly exhaustive of all possibilities, then their probabilities given the same condition C add up to 1 (or to 100%). For example, the probability that Albert was at the crime scene given that Carole says she saw him and the probability that Albert was not at the crime scene given that Carole says she saw him add up to 1 (or to 100%). So if we estimate the first of these probabilities as 0.75 (or 75%) then the latter probability is 0.25 (or 25%).

An important concept in probability theory is statistical independence. Two statements A and B are *statistically independent* of each other if information about the probability that A is true is irrelevant for determining the probability that B is true, otherwise they are *statistically dependent*. For instance, let A be ‘I flip a coin the first time’ and B ‘I flip the coin a second time’. Assuming the coin is fair, the probability that I flip heads the first time is 0.5 (that is, 50%). Clearly, the probability that I flip heads the second time also is 0.5 (50%) since the two coin-flipping events do not influence each other. When two statements A and B are statistically independent, then the probability that they are both true (‘A and B’) can be calculated by multiplying the

² Parts of this subsection are taken from Prakken (2026).

probabilities that each of them is true. So the probability that I flip heads both the first and the second time equals $0.5 \times 0.5 = 0.25$ (or 25%). However, when A and B are statistically dependent of each other, then this calculation is invalid and the probability of A and B has to be determined independently from the probabilities of A and B.

Conditional probabilities allow a precise definition of when two statements A and B are statistically independent of each other. We say that statement A is statistically independent of statement B if the conditional probability of A given B equals the unconditional probability of A. In other words, whether B is true or not does not matter for the probability that A is true. This can be generalised to statistical(in)dependence of conditional probabilities: statement A is statistically independent of statement B conditionally on statement C if the conditional probability of A given B & C equals the conditional probability of A given C. The multiplication rule applies here too: the conditional probabilities of A given C and B given C can be multiplied to give the conditional probability of A & B given C if and only if A and B are statistically independent of each other given C.

One way to determine conditional probabilities is to directly verify them in a population. For instance, if 70% of Dutch adults is married, we can say that the conditional probability of a given Dutch person is married given that he or she is an adult is 70%. Another way to determine a conditional probability is to calculate it with so-called Bayes' theorem from other probabilities of a required type. For instance, if we know the probability that an arbitrary Dutch person is married and if we know the probability that an arbitrary Dutch person is an adult, respectively, not an adult given that s/he is married, then we can calculate the probability that the person is married given that s/he is a Dutch adult. For present purposes we do not need to specify the precise form of Bayes' theorem. See e.g. Hacking (2001) for an excellent introduction and e.g. Prakken (2026) for legal applications.

It is sometimes said that the use of Bayes' theorem commits to the view that the involved probabilities are subjective in that they would just describe a person's subjective degree of belief in a proposition (e.g. Hahn & Oaksford 2006, pp. 8-9). However, this is not true in general. Philosophers distinguish *frequency-type* and *belief-type* probability. Probabilities are of the frequency type if they are based on relative frequencies, resulting, for instance, from experiments that can be repeated indefinitely, such as tossing a coin or rolling a dice, or from observing distributions in a population, such as checking how many Dutch adults are married. In contrast, probabilities are of the belief type if they are about the degree to which a proposition is believable. Such probabilities can be attached to generalizations (e.g. 'The probability that any given Dutch adult is married is 70%') but they can also be attached to propositions that a single event occurs (e.g. 'the probability that Dutch adult John is married is 80%'). In sum, probabilities about classes of events or states of affair can be both frequency-type and belief type while probabilities about individual events or states of affair are always belief-type. Both types of probabilities can be more or less well-founded, so they can both be more or less objective or subjective. Most importantly, both types of probabilities are subject to the same laws of probability theory, including Bayes' theorem. See for all this Chapter 11 of Hacking (2001).

2.2 ASPIC+: formalising the structure of arguments and the nature of attack and defeat³

ASPIC+ is a formal framework for specifying the structure of arguments and the nature of their relations of attack and defeat. The ideas underlying ASPIC+ are that reasoning consists of constructing arguments by applying inference rules, by attacking these arguments with counterarguments, and by adjudicating between conflicting arguments on grounds that are appropriate to the conflict at hand. Just as in deductive reasoning, arguments must instantiate inference rules (which can be regarded as argument schemes) but only some of these rules capture fool-proof reasoning. ASPIC+ assumes two types of inference rules: the strict ones, of which the truth of all their antecedents guarantees the truth of their conclusions, and the defeasible ones, of which the truth of all their antecedents only creates a presumption of the truth of their conclusions.

ASPIC+ can be instantiated in many different ways. For example, the strict rules can be chosen to be the rules of some deductive logic, and the defeasible rules can be chosen to correspond to some adopted set of presumptive argument schemes (Walton 1996). However, for present purposes we can keep things simple. I assume that all premises are certain and that the logical language in which an argument's premises and conclusion can be expressed consists of so-called propositional-logic literals. That is, they are either atomic propositions or their negations. Then each step in an ASPIC+ argument is of the form

$$P_1, \dots, P_n, \text{ therefore } C$$

satisfying the following conditions.

- Each P_i and C is a propositional atom or a negation thereof.
- The inference step is supposed to be based on a generalisation *if* $P_1, \dots, P_n$ *then* (*qualifier*) C , where the qualifier is an expression of strength of the generalization. A generalisation is *defeasible* if the qualifier is something like 'normally', 'typically', 'usually', 'most of the time' and *strict* if the qualifier is something like 'always' or 'certainly'. Generalisations can be seen as the warrants in Toulmin's (1958) famous argument scheme.
- Each P_i is either given as evidence or a conclusion from a preceding inference in the argument.

Arguments are then trees of such inference steps, where the leaves are its *premises* (pieces of evidence), the links are inference steps, the root is the *final conclusion* and all other nodes are *intermediate conclusions*. Subtrees of an argument are *subarguments* of the argument. In my 2018b paper I further assume that the set of all conclusions of an argument is consistent and that different subarguments of an argument must have different conclusions (which, among other things, excludes circular arguments).

Figure 1 displays an argument for the final conclusion c_3 that a suspect S committed some crime, with three premises p_1 , p_2 and p_3 and two subarguments with intermediate conclusions c_1 and c_2 .

³ This informal summary of ASPIC+ is largely taken from Prakken (2020).

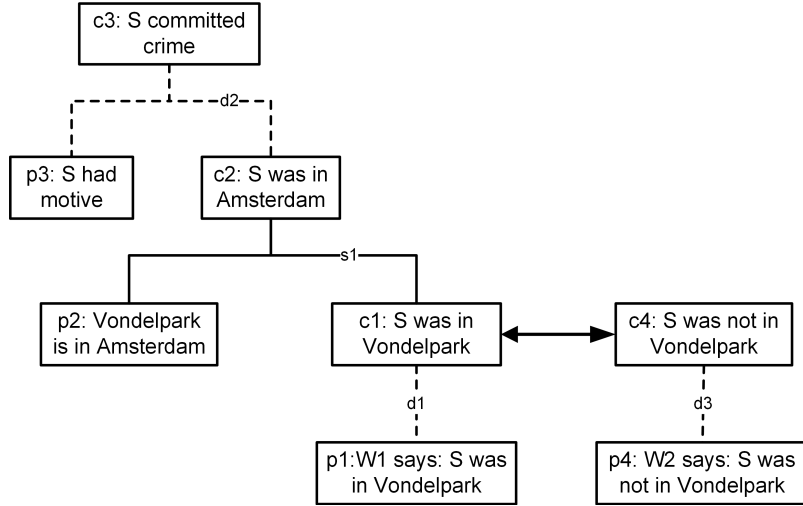


Figure 1: An argument and a counterargument

In the simple ASPIC+ instantiation we are considering, arguments can be *attacked* in just one way, namely, with an argument of which the conclusion contradicts one of their defeasibly derived conclusions (*rebutting attack*). Rebutting attack can be either *direct* (contradicting the final conclusion) or *indirect* (contradicting an intermediate conclusion).

Figure 1 displays one attacking counterargument, viz. the argument with final conclusion c3 that suspect S was not in the Vondelpark. This argument directly rebuts the argument with conclusion c1 and it indirectly rebuts the arguments with conclusions c2 and c3.

ASPIC+ then distinguishes between *attack* and *defeat* relations between arguments. Attacks succeed as defeats if the attacker is not weaker than it target according to some preference criterion. Then argument A *defeats* argument B *on its subargument* B' if A directly rebuts B' and A is at least as strong as B'. And we say that A *defeats* B if A directly defeats B on one of B's subarguments. Note that this leaves open the possibility that two attacking arguments defeat each other, namely, if they are equally strong. If A defeats B but B does not defeat A, we say that A *strictly defeats* B.

In Figure 1 the arguments that have to be compared on their relative strength are the arguments for the conclusions c1 and c4. Assume for the sake of illustration that latter is stronger than the former. Then the latter directly defeats the former and indirectly defeats the arguments for conclusions c2 and c3.

2.3 The theory of abstract argumentation frameworks

That an argument A strictly defeats an argument B does not yet resolve the conflict, since A may in turn be (strictly or weakly) defeated by an argument C, and so on. In some cases the outcome can still be easily determined. For example, a common pattern, depicted in Figure 2a below, is that A is strictly defeated by B which is strictly defeated by C. In this case it is clear that C defends A against B so both A and C can be accepted while B must be rejected. To illustrate this with Figure 1, suppose that premise p4 is not directly given but is itself inferred from a premise p5 that some police document says that W2 says that S was not in the Vondelpark, and imagine that this argument is strictly

defeated by some argument with as conclusion that W2 did not say that S was in the Vondelpark. Then we have an example fitting the format of Figure 2a.⁴

However, we can easily imagine more complex examples where our intuitions fall short. For instance, another argument D could be constructed such that C and D defeat each other, then an argument E could be constructed that defeats D but is defeated by A, and so on: which arguments can now be accepted and which should be rejected? Here we cannot rely on intuitions but need a calculus. Its input will be a collection of arguments plus an assessment of which arguments defeat each other, while its output will be an assessment of the *contextual status* of these arguments in terms of three classes (three and not two since some conflicts cannot be resolved). Intuitively, the *justified* arguments are those that survive all conflicts with their attackers and so can be accepted, the *overruled* arguments are those that are attacked by a justified argument and so must be rejected; and the *defensible* arguments are those that are involved in conflicts that cannot be resolved.

In artificial intelligence the standard calculus for determining the dialectical status of arguments is Dung's (1995) theory of so-called abstract argumentation frameworks. This theory says nothing about the structure of arguments or the nature of defeat. I now explain the basic ideas in a labelling-based form, which is not in Dung's original paper but was later proved to be equivalent to Dung's account (see Baroni et al. 2011).

The idea is to label all arguments in the tree as *in* or *out* according to the following constraints:

1. An argument is *in* if and only if all its defeaters are *out*
2. An argument is *out* if and only if it has a defeater that is *in*

The application of these two rules can be surprisingly subtle and complex (see e.g. Prakken 2018a) but in the present paper we will consider just a few types of simple argumentation frameworks and we will only focus on labellings that maximise the set of arguments labelled *in*. Consider the following two example frameworks, in which the nodes are arguments and the links are defeat relations:

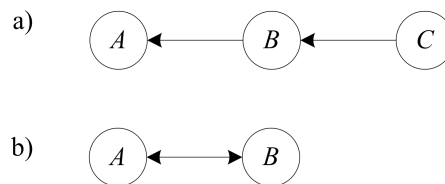


Figure 2: two abstract argumentation frameworks

In Figure 2a there is a unique maximal labelling, in which A and C are *in* and B is *out*: A has to be *in* by constraint (1) since it has no defeaters, so trivially all its defeaters are *out*. Then B must be *out* since it has a defeater that is *in*. Finally, C must be *in* since all its defeaters (in this case B) are *out*.

⁴ Strictly speaking there are more arguments in the example, since the given arguments have subarguments. However, since they do not matter for the labelling of arguments A, B and C, we leave these arguments implicit.

In Figure 2b things are more subtle. If we label A *in* and B *out*, then both constraints are satisfied, so A defends itself against B. Conversely, if we label A *out* and B *in*, then both constraints are also satisfied, so B defends itself against A.

Accordingly, we define an argument as *justified* if it is labelled *in* in all labellings, as *overruled* if it is labelled *out* in all labellings, and *defensible* otherwise. So in Figure 2a, both A and C are justified while B is overruled, and in Figure 2b both A and B are defensible.

To illustrate this with the example of Figure 1, let us focus on the three arguments A1 with conclusion c1, A3 with conclusion c3 and B with conclusion c4 (their subarguments do not matter for the labellings in this example so we leave them implicit). See Figure 3.

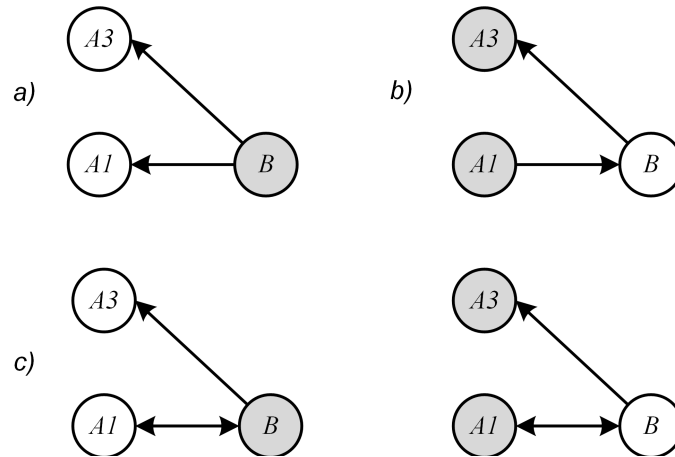


Figure 3: AFs and labellings corresponding to Figure 1

If B strictly defeats A1 then B also strictly defeats A3, so we have a unique labelling in which B is *in* and A1 and A3 are *out*. This is depicted in Figure 3a: the grey colour stands for *in* and the white colour stands for *out*. So B is justified while A1 and A3 are overruled. If instead A1 strictly defeats B then we have a unique labelling in which B is *out* and A1 and A3 are *in*, (depicted in Figure 3b) so B is overruled while A1 and A3 are justified. Finally, if A1 and B defeat each other then B also strictly defeats A3 and we have two labellings, namely, the ones just given for the first two cases (depicted, respectively, on the left and right in Figure 3c). Then all three arguments are defensible.

3 Logical, dialectical and rhetorical argument strength⁵

Before showing how ASPIC+ arguments can be given probabilistic strength, it is useful to discuss what kind of argument strength is thus modelled. Here it is natural to take Aristotle's famous distinction between logic, dialectic and rhetoric as starting point. Very briefly, *logic* concerns the validity of arguments given their form, *dialectic* is the art of testing ideas through critical discussion and *rhetoric* deals with the principles of effective persuasion (Van Eemeren et al. 2014, Section 1.4). Accordingly, I distinguish between logical, dialectical and rhetorical argument strength.

⁵ Parts of this section have been summarized from Section 3 of Prakken (2021).

Logical argument strength divides into two aspects: inferential and contextual argument strength.

Inferential argument strength is about how well the premises support the conclusion if we only look at the argument's premises, inferences and conclusion(s). For instance, arguments with only deductive inferences are inferentially stronger than arguments with defeasible inferences. Our notion of probabilistic argument strength will be of this kind.

Contextual argument strength is about how well the conclusion of an argument is supported if we look at the context of all relevant arguments in a given state of a discussion or inquiry. This is the kind of argument strength modelled by Dung's theory of abstract argumentation frameworks. The reader might wonder why this is not called dialectical strength, since after all, determining an argument's contextual strength as defined here involves the comparison of argument and counterargument. Yet this is not truly dialectical, since the arguments are all assumed to be constructed from a given body of information and their structure and relations are determined by syntactic and semantic means.

Dialectical argument strength looks at how well defended an argument is in the context of an ongoing or terminated critical discussion. Zenker et al. (2020) formulate this as "a function of the (un)availability of permissible move sequences originating at the present dialogue stage, and ending in a discussant's role-specified goal being achieved". For example, one might regard an argument as dialectically weaker the more attacks on it are possible in the current state. This underlies an idea that many decision makers are aware of, namely, to justify one's decisions as sparsely as possible, in order to minimize the chance of successful appeal. The main difference between contextual and dialectical argument strength is that the former is static, determined by all arguments given at a particular state, while the latter is dynamic, considering how the contextual status of an argument can change by adding new arguments.

Finally, rhetorical argument strength looks at how capable an argument is to persuade other participants in a discussion or an audience (Perelman & Olbrechts-Tyteca 1969). Persuasiveness is thus essentially a psychological notion.

4 Probabilistic argument strength in ASPIC+

In this section I will explain how ASPIC+ arguments have been given probabilistic strengths in Prakken (2018b) and I will discuss how this relates to the Bayesian approaches to argument strength in argumentation theory and psychology. Before doing so, it is important to discuss what kinds of arguments can have probabilistic strength. At first sight, it might be thought that arguments about what is the case can have this, for instance, arguments that apply empirical generalizations, as in *Birds can typically fly, Tweety is a bird, therefore (presumably), Tweety can fly*. However, several argumentation-theorists who have applied a Bayesian approach to argument strength have conducted experiments on degrees of belief that humans attach to the premises and conclusions of arguments (e.g. Oaksford & Hahn 2013). According to the above distinction between three kinds of argument strength made in Section 3, this strictly speaking addresses rhetorical strength, namely, how persuasive an argument is for an audience. Yet it might also be said that such studies equate rhetorical and logical argument strength, since one aim of such studies is to check whether humans assign their degrees of belief in accordance with Bayes' theorem, which these studies regard as

a norm for sound argumentation. Arguably, degrees of belief can also be assigned to, for instance, legal, moral and political arguments. For this reason I will not commit myself to any claim as to which arguments are and which are not amenable to having probabilistic strength. Even though all my examples will involve generalisations, I believe that my analysis in this section applies to all types of arguments that can reasonably be argued to have probabilistic strength in whatever sense of probability, as long as it conforms to the laws of probability theory.

It is also relevant to discuss an often-made objection to probabilistic approaches, namely, that the required probabilities to make such approaches work are often not available in practice. Granted that the latter is true, this does not make the theoretical study of probabilistic approaches practically useless. One reason for this is that the fallaciousness of various types of qualitative inferences can be explained in terms of probability theory. One example is the deductive invalidity of the contraposition of statistical generalisations. That most British people do not speak Welsh does not imply that most Welsh-speaking people are not British; in fact, the great majority of Welsh-speaking people are British. So the theoretical study of probabilistic approaches can be practically relevant even if precise probabilities are not always available in practice.

4.1 Probabilistic inferential strength of multi-steps arguments⁶

The first idea then of Prakken (2018b) is to assume that all inference rules have a probabilistic strength that expresses the conditional probability of the rule's consequent C given the conjunction of its antecedents $P_1, \dots, P_n$. They can thus be seen as generalisations. Defeasible inference rules have a strength greater than 0.5 but less than 1, while strict rules have strength 1. The constraint that defeasible rules have strength greater than 0.5 (taken from Pollock 1995) agrees with the intuitive reading of defeasible rules as *If P then usually/typically/normally C* . Because of this reading, we do not allow that both a rule *If $P_1, \dots, P_n$ then C* and a rule *If $P_1, \dots, P_n$ then not- C* are available.

The second idea is to use these rule strengths to define a probabilistic notion of inferential argument strength (in Prakken 2018b called an argument's internal strength). A main point here is that since arguments can have internal structure, this notion cannot in general be defined as the conditional probability of an argument's conclusion given the conjunction of its premises. Consider the following example:

Bart is Dutch, so Bart likes cycling, so Bart is healthy
Bart is Dutch, so Bart likes swimming, so Bart is healthy

If argument strength equates the probability of its conclusion given its premises, then both arguments are equally strong, since they have the same premise and the same conclusion. However, if the rules applied in these arguments are defeasible, they may well have different strengths, since the health statistics may be different for swimmers than for cyclists, or people may assign different degrees of belief to them.

The main idea of Prakken (2018b) is that this limitation can be overcome by defining argument strength as the probability of the conjunction of all premises and conclusions of an argument. For present purposes, since we assume the premises to be given as being certain, this equals the conditional probability of the conjunction of all

⁶ The text of this and the next subsection is largely taken from Prakken (2018b).

conclusions of an argument given the conjunction of its premises.⁷ This is the definition that will be investigated in this paper. According to probability theory this notion of strength can be calculated with the general version of the chain rule for probability distributions. Consider any argument A with premises and conclusions $C_1, \dots, C_n$, and consider any top-down breadth-first order of the argument's conclusions and premises, starting with its final conclusion C_n , when regarding an argument as an inference tree with as nodes the premises and conclusions of A and as links any application of an inference rule in A . Then the general chain rule amounts to the following equation, in which $\Pr(Y | X)$ stands for the conditional probability of Y given X .

$$\Pr(C_1 \& \dots \& C_n) = \Pr(C_n | C_1 \& \dots \& C_{n-1}) \times \Pr(C_{n-1} | C_1 \& \dots \& C_{n-2}) \times \dots \times \Pr(C_1).$$

However, we are not yet there, since a practical problem with the chain rule definition of argument strength is that in many cases the conditional probabilities needed for this calculation will not be available. To see this, consider the argument

I see smoke, my observations are usually correct, therefore there is smoke. Where there is smoke, there usually is fire, therefore there is fire

abstracted to

A, therefore, presumably, B, therefore, presumably, C

The chain rule implies that $\Pr(C \& B \& A)$ equals $\Pr(C | A \& B) \times \Pr(B|A) \times \Pr(A)$. So we have to know the conditional probability that there is fire (C) given that I see smoke (A) and that there is smoke (B). However, the arguments were constructed with the rules *If A then usually B* and *If B then usually C*, so the only probabilities that are given as rule strengths are $\Pr(C|B)$ and $\Pr(B|A)$ (as well as $\Pr(A) = 1$). This will arguably often be the case in practice, where people often specify a rule *If X then normally Y* with strength $\Pr(Y|X)$ without also specifying a more specific rule *If X and Z then normally Y* with strength $\Pr(Y | X \& Z)$. In our smoke and fire example, the only probabilities that will usually be given are the probabilities that there is smoke given that I see smoke, i.e., $\Pr(B|A)$ and that there is fire given that there is smoke, i.e., $\Pr(C|B)$. So in practice the general chain rule will often not be applicable since the required probabilities will not be available.

To illustrate the same with the example of Figure 1, the above chain rule yields:

$$\Pr(c_1 \& c_2 \& c_3) = \Pr(c_3 | c_1 \& c_2 \& p_1 \& p_2 \& p_3) \times \Pr(c_2 | c_1 \& p_1 \& p_2 \& p_3) \times \Pr(c_1 | p_1 \& p_2 \& p_3)$$

(strictly speaking this should be continued with $\times \Pr(p_3 | p_1 \& p_2) \times \Pr(p_2 | p_1) \times \Pr(p_1)$ but since the premises are assumed to be certain, these amount to multiplying by 1).

Fortunately, there is a way out of this problem. In many practical contexts, agents will specify a rule *If X then normally Y* since, as far as they know, X is all that is needed to conclude Y . If they had thought that Z is also needed to conclude Y , they would

⁷ It is straightforward to generalise my approach to the case with possibly uncertain premises but for simplicity I will in this article assume that all premises are certain.

instead have specified a rule *If X and Z then normally Y*. More generally, it seems reasonable to assume that human reasoning operates under the general assumption that the conclusion of an inference step in an argument is, conditionally on its premises (the antecedents of the applied rule), statistically independent of anything else. In our abstract example, this assumption amounts to saying that $\Pr(C | A \& B)$ equals $\Pr(C|B)$, which allows us to apply the chain rule. Now according to probability theory this assumption, together with our assumption that all premises are certain, amounts to saying that the general chain rule is equivalent to the following special version (where $\text{ant}(C_i)$ denotes the set of antecedents of the rule in the argument that yields conclusion C_i).

$$\Pr(C_1 \& \dots \& C_n) = \Pr(C_1 | \text{ant}(C_1)) \times \dots \times \Pr(C_n | \text{ant}(C_n)).$$

This equation says that the probabilistic strength of an argument can be calculated by multiplying all rule strengths of any rule applied in the argument.

In our example of Figure 1 this yields:

$$\Pr(c_1 \& c_2 \& c_3) = s(d1) \times s(d2) \times s(d3) \times s(s1) = s(d1) \times s(d2) \times s(d3)$$

which is a much simpler calculation than the one induced by the general chain rule. For example, if $s(d1) = 0.8$ and $s(d2) = 0.9$ and $s(d3) = 0.7$, then the strength of the argument for $c3$ is $0.8 \times 0.9 \times 0.7 = 0.504$.

This example also shows that an argument's strength can be less than 0.5. For instance, if $s(d2)$ is changed to 0.8, then the strength of the argument for $c3$ is $0.8 \times 0.8 \times 0.7 = 0.448$. At first sight, this would seem to be a weakness of our approach. However, note that an argument strength of less than 0.5 does not imply that the probability of its conclusion is less than 0.5: it will at least be 0.448 and possibly higher. Moreover, in practical applications an argument can be very uncertain even if no attacker can be constructed. This can happen, for instance, if the argument chains several defeasible rules. In such cases the probabilistic strength of an argument gives useful additional information to an argument evaluator. For example, the evaluator might in the end decide to not accept arguments that have strength less than 0.5 even if they are justified according to the argumentation logic.

Completing our discussion so far, it should be noted that by giving rules a probabilistic strength, we have in fact modelled what Pollock (1995) called the *statistical syllogism*, which says that if P holds and the probability of C given P is x, then C may be inferred from P with strength x. Pollock then defined the strength of multi-steps arguments as the minimum of the strengths of the individual inference steps in the argument (his weakest-link principle). In our example, the strength of the argument for $c1$ would thus be 0.7. In Prakken (2018b) I proved that the above definition with the special chain rule is equivalent to the weakest link principle if an argument applies at most one defeasible rule but not in general. (This result shows that the approach of my 2018b paper avoids the criticism of Hahn & Oaks (2006) of weakest-link approaches).

4.2 Probabilistic argument strength in the light of counterarguments

The above is not yet all there is to say, since the independence assumption that justifies the use of the reduced chain rule is clearly invalid in general. Consider the following example.

People who live in Denmark (D) usually speak Danish (S) but English-speaking university employees (E) who live in Denmark usually do not speak Danish.

The implicit assumption of an argument using the first rule that $\Pr(S|D) = \Pr(S | D \& E)$ is contradicted by the second rule strength, since all rules are assumed to have strength greater than 0.5 while the axioms of probability theory imply that $\Pr(S | D \& E) + \Pr(\text{not-}S | D \& E) = 1$. Nevertheless, it seems reasonable to say that the assumption holds in the absence of information to the contrary. On this account, the independence assumption is defeasible and its defeasibility is captured by the possibility of an attacking argument. In our example we have, for a given English-speaking academic who lives in Denmark, the following two arguments which rebut each other.

A1 = D therefore S

A2 = D, E, therefore not-S

A variant of this example is the following example:

Adults are usually married but students are usually adults and students are usually not married.

For a given student this yields the following arguments:

A3 = Student, therefore Adult, therefore Married

A4 = Student, therefore, not Married

Argument A3 assumes that $\Pr(\text{Married}|\text{Adult}) = \Pr(\text{Married} | \text{Adult} \& \text{Student})$ while Argument A4 assumes that $\Pr(\text{Not-Married}|\text{Student}) = \Pr(\text{Not-Married} | \text{Adult} \& \text{Student})$. Again, since all rules are assumed to have strength greater than 0.5, these two assumptions are jointly inconsistent.

The upshot of all this is that resolving conflicts between defeasible arguments by just resorting to their internal inferential argument strengths is problematic, since the arguments assume different and jointly inconsistent probability distributions. This also implies that the proposed notion of probabilistic argument models internal argument strength (see Section 3 above). To illustrate this, consider again the speaking-Danish example and assume that 90% of the people who live in Denmark speak Danish while 75% of the English-speaking academics who live in Denmark does not speak Danish. Do we then want to accept argument A1 that our English-speaking academic who lives in Denmark presumably speaks Danish? Of course not, since the second statistic is about a more specific class than the first one, so the generally accepted principle to adopt the statistic about the more specific reference class requires that we instead conclude that our person presumably does not speak Danish. For this reason, probability-based comparisons between arguments should, either explicitly or implicitly, involve a kind of specificity principle, which implies that conflicts between defeasible arguments cannot in general be resolved by just resorting

to their probabilistic argument strengths. For example, Pollock's (1995) so-called subproperty defeater of statistical inferences is such a specificity principle.

These observations can be extended to cases in which the attacking arguments do not have a specificity relation. Consider the following well-known example from nonmonotonic logic:

Quakers are usually pacifists, Republicans are usually not pacifists, Nixon was a quaker and a republican.

This gives rise to the following rebutting arguments.

A5: Nixon was a quaker, therefore Nixon was a pacifist

A6: Nixon was a republican, therefore Nixon was not a pacifist

Here argument A5 implicitly assumes that $\Pr(\text{Pacifist} | \text{Quaker}) = \Pr(\text{Pacifist} | \text{Quaker} \ \& \ \text{Republican})$ while argument A6 implicitly assumes that $\Pr(\text{not-Pacifist} | \text{Republican}) = \Pr(\text{not-Pacifist} | \text{Quaker} \ \& \ \text{Republican})$, which again is inconsistent given that rule strengths exceed 0.5. Suppose, furthermore, that $\Pr(\text{Pacifist} | \text{Quaker}) > \Pr(\text{not-Pacifist} | \text{Republican})$. Are we then forced to accept argument A5? No, since what we want to know is $\Pr(\text{Pacifist} | \text{Quaker} \ \& \ \text{Republican})$ and this probability may be assumed independent of the former probabilities. So in the absence of information about $\Pr(\text{not-Pacifist} | \text{Quaker} \ \& \ \text{Republican})$ it seems just as rational to regard both arguments as defensible as preferring A5 over A6.

Our analysis suggests that conflict resolution in the setting of probabilistic argumentation as modelled in Prakken (2018b) may be seen as the *adjustment* of probabilities to obtain a consistent probability distribution that can be used to resolve conflicts between arguments. In my 2018b paper I explore some possible approaches that do this but in the present paper I confine myself to listing some properties that are proven in my 2018b paper about the above method. First, for any set of arguments that is not conflict-free in that it contains defeating arguments it can be shown that their joint inferential strengths are, when combined with the special chain rule, inconsistent with the axioms of probability. Second, the converse does not hold in general. Consider two defeasible rules *If X then normally Y* and *If X and Z then normally Y*. If these rules have different rule strengths and if both X and Z are given, then this yields two different arguments for Y. These arguments do not defeat each other but their rule strengths are, when joined with the special chain rule, inconsistent. However, the following result can be shown on three assumptions: that the set of all premises of any argument (the evidence) is consistent, that there cannot be more than one rule with the same consequent, and that for every strict rule its contraposition is also available. Given these assumptions it holds for every set of arguments that are labelled *in* in the same labelling that their joint inferential strengths are, when combined with the special chain rules, consistent with the axioms of probability theory.

4.3 Argumentation and Bayesian reasoning

A proponent of Bayesian approaches might at this point conclude that, if probabilistic inferential argument strength cannot be used to resolve conflicts between arguments, it is better to resort to a fully probabilistic approach. This is indeed what Hahn et al.

(2013) do with the argument scheme of appeal to expert opinion which they encode in a so-called Bayesian network, including the scheme's critical questions. Bayesian networks (BNs) are the state-of-the art in probabilistic reasoning. They use graph theory to graphically represent statistical independence relations. Formally, a BN is a directed acyclic graph in which the nodes stand for probabilistic variables, or propositions that can be true or false⁸ and in which a link from one node to another indicates that the truth value of the former possibly statistically depends on the truth value of the latter. Despite their name, BNs do much more than just applying Bayes' theorem; in fact, they support the calculation of any probability of interest concerning the probabilistic variables they contain. Interestingly, the special chain rule used above to compute an argument's inferential strength can also be used to compute the probability of any truth-value assignment to all variables in a BN. So, formally speaking, in our approach of Section 4.1 arguments are mini BNs with partially specified conditional probability tables (see below for what that means).

Figure 4 (taken from Timmer et al. 2017) displays a small BN concerning a criminal case. The conditional probability tables are shown as tables inside the nodes of the graph.

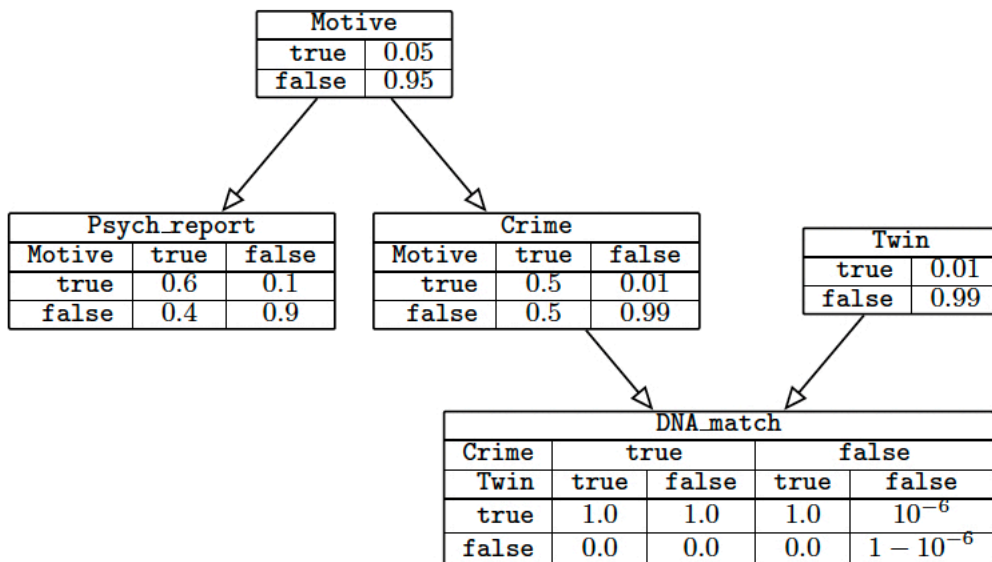


Figure 4: An example Bayesian Network

The variable Crime stands for whether the suspect committed the crime, Motive stands for whether the suspect had a motive to commit the crime, Psych_report captures whether the psychologist reported that the suspect has a motive and Twin stands for whether the suspect has a twin brother. Finally, DNA_match stands for whether the suspect's DNA matches with DNA found at the crime scene. The arrows represent possible statistical dependencies. Hence, each node has a so-called conditional probability table (CPT), which specifies for each combination of truth values of all its parents what is the probability that the node is true, respectively, false. For example, the

⁸ They can in general have multiple variables but for simplicity we only consider BNs with two-valued variables.

CPT of Crime says that if Motive = true, then the probability that Crime = true is 0.5 (and so the probability that Crime = false is 0.5), while if Motive = false, then the probability that Crime = true is 0.1 (and so the probability that Crime = false is 0.9). Moreover, the CPT of DNA-match says, for instance, that if Crime = false then if Twin = true then the probability of DNA_match = true is (almost) 1 (so the probability of DNA_match = false is almost 0). When a node has no parents, its CPT captures the unconditional probability that it is true, respectively, false.

Consider next the following argument and counterargument, which loosely corresponds to the information in the BN.

The psychologist reported that the suspect has a motive, therefore the suspect has a motive. Moreover, the suspect's DNA matches with DNA found at the crime scene. Therefore, the suspect committed the crime.

The suspect has a twin brother, the twin brother's DNA matches with the DNA found at the crime scene, therefore, the suspect's twin brother committed the crime, therefore, the suspect did not commit the crime.

We could directly model these arguments in ASPIC+ with the method of Section 4.1 or we could enter the three pieces of evidence Psych_report = true, Twin = true and DNA_match = true into the BN and compute the conditional probability of Crime = true given these three pieces of evidence. If we follow the argumentation approach proposed in this paper, then the first thing to note is that we cannot use the probabilities in the CPT of Figure 4 as rule strengths, since the BN was (as usually with BN design) designed to capture causal relations, while the argument (as often in argumentation) reasons with evidential relations. For instance, the BN says that the suspect having a motive can cause the psychologist to say that the suspect had a motive, while the argument infers from the psychologist's testimony that the suspect had a motive. This shows that moving from an argument-based to a BN approach is not straightforward: it often also commits to a different style of modelling the relevant knowledge, namely, from cause to effect instead of from effect to cause (See Pearl 1988 and Prakken 2014 for more on the difference between causal and evidential probabilistic reasoning).

The second thing to note if we follow our argumentation approach is that we end up with two rebutting arguments with jointly inconsistent probabilistic strengths. To resolve this conflict, we would have to adjust these strengths to make them consistent. By contrast, the BN simply computes the probability of interest from the given BN and the evidence, without the need for any adjustment. Moreover, it does so according to the laws of probability theory.

For these reasons it might be thought that a full BN approach is theoretically superior to this paper's argumentation-based approach and that examples with conflicting arguments have to be modelled with BNs. However, there are reasons to doubt this and to regard argumentation-based approaches as an interesting alternative. Granted that if all required probabilities are available and are jointly consistent, the case for the theoretical superiority of a full BN approach is strong, the point is that this ideal situation is not always the case. Above we have seen the importance of defeasible independence assumptions and conflict resolution but BNs cannot handle either: the independence assumptions are hardwired into a BN's structure, and the way it calculates probabilities assume that the specified probabilities are jointly consistent. By contrast,

these are strong points of an argumentation-based approach, since it can model defeasible assumptions and conflict resolution. The first is desirable since we have seen that without the special chain rule and its underlying defeasible independence assumptions it seems hard to rationally justify the chaining of defeasible rules in arguments, as happens in our above examples and as arguably happens all the time in everyday human reasoning.⁹ And the means to resolve conflicts is desirable since the information that humans have available is not always consistent, so humans often organize their reasoning in terms of explicit arguments and counterarguments. Among other things, the idea (explored in Prakken 2018b and briefly mentioned above at the end of Section 4.2) that people, when faced with conflicting information, adjust their (degrees of) belief to make them consistent, is a natural one. So there is every reason to study argumentation-based approaches to reasoning with uncertain information, even if numeric probabilities are available. This is not to say that argumentation approaches are superior to purely Bayesian approaches. All I am claiming is that in context where a single probability distribution is not obviously available, argumentation-based approaches are an interesting alternative, or at least an interesting addition to purely Bayesian approaches.

5 Conclusion

In this paper I have summarised my ideas in Prakken (2018b) on defining a probabilistic notion of argument strength in the context of logics of argumentation, and I have discussed the relevance of my proposal to Bayesian accounts of argument strength in argumentation theory and philosophy. My 2018b account is based on the ideas that the probability of an argument is the probability of the conjunction of all its premises and conclusions and that arguments implicitly make probabilistic independence assumptions. Together these ideas imply that the probabilistic assumptions of conflicting arguments are jointly inconsistent. This in turn implies that resolving conflicts between arguments in terms of their internal strengths is problematic.

On the basis of this I have argued that my 2018b approach reveals two challenges for the Bayesian approach to argument strength: multi-steps arguments highlight the need for defeasible probabilistic-independence assumptions and the possibility of conflicting arguments highlights the need for conflict resolution. I have argued that the modelling of both phenomena goes beyond standard probability theory, even if it is cast in the form of Bayesian networks. I have shown that both phenomena can, at least to some extent be modelled in an argumentation logic in a way that makes use of but goes beyond probability theory. There may be other ways of doing so, but at least my paper has shown how far we can get in an approach in which arguments and counterarguments are first-class citizens.

References

⁹ The idea that reasoning under uncertainty involves defeasible probabilistic assumptions is not new. For instance, epsilon semantics (Goldszmidt et al. 1993), a model-theoretic approach to formalizing reasoning with defeasible conditionals in agreement with probability theory but without specifying numbers, makes a maximum-entry assumption on probability distributions, which informally stands for the assumption that the distributions are as regular as possible. This defeasibly validates, for instance, contraposition and chaining of statistical generalisations.

- Baroni, P., Caminada, M. & Giacomin, M. (2011), An introduction to argumentation semantics. *The Knowledge Engineering Review* 26(4), 365-410.
- Dung, P.M. (1995), On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n -person games. *Artificial Intelligence*, 77, 321-357.
- Eemeren, F. van, Garssen, B., Krabbe, E., Henkemaans, A.S., Verheij, B. & Wagemans, J. (2014), *Handbook of Argumentation Theory*. Springer, Dordrecht.
- Goldschmidt, M., Morris, P. & Pearl, J. (1993), A maximum entropy approach to nonmonotonic reasoning. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(3), 220-232.
- Hacking, I. (2001), *An Introduction to Probability and Inductive Logic*. Cambridge University Press, Cambridge.
- Hahn, U. & Hornikx, J. (2016), A normative framework for argument quality: Argumentation schemes with a Bayesian foundation. *Synthese* 193, 1833-1873.
- Hahn, U. & Oaksford, M. (2006), A normative theory of argument strength. *Informal Logic* 26 (1), 1-24.
- Hahn, U., Oaksford, M. & Harris, A.J.L. (2013), Testimony and argument: a Bayesian perspective. In Zenker (2013), pp. 15-38.
- Hunter, A.J., Polberg S., Potyka, N., Rienstra, T. & Thimm, M. (2021), Probabilistic argumentation: a survey. In D. Gabbay, M. Giacomin, G. Simari & M. Thimm (eds.): *Handbook of Formal Argumentation*, Vol. 2, pp. 397-444. College Publications, London.
- Modgil, S.J. & Prakken, H. (2014), The *ASPIC+* framework for structured argumentation: a tutorial. *Argument and Computation* 5, 31-62.
- Modgil, S.J. & Prakken, H. (2018), Abstract rule-based argumentation. In P. Baroni, D. Gabbay, M. Giacomin & L. van der Torre (eds.): *Handbook of Formal Argumentation*, Vol. 1, pp. 286-361. College Publications, London.
- Oaksford, M., & Hahn, U. (2013), Why are we convinced by the Ad Hominem argument? Bayesian source reliability or pragma-dialectical discussion rules. In Zenker (2013), pp. 39-58.
- Pearl, J. (1988), Embracing causality in default reasoning. *Artificial Intelligence* 35, 259-271.
- Perelman, C. & Olbrechts-Tyteca, L. (1969), *The New Rhetoric. A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame, Indiana.
- Pollock, J.L. (1995), *Cognitive Carpentry. A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA.
- Prakken, H. (2010), An abstract framework for argumentation with structured arguments. *Argument and Computation* 1, 93-124.
- Prakken, H. (2014), On direct and indirect probabilistic reasoning in legal proof. *Law, Probability & Risk* 13, 327-337.

- Prakken, H. (2018a), Argument. In S.O. Hanson & V.F. Hendricks (eds.): *Introduction to Formal Philosophy*, pp. 63-79. Springer Undergraduate Texts in Philosophy, Springer International Publishing AG.
- H. Prakken (2018b), Probabilistic strength of arguments with structure. In *Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018)*, pp. 158-167.
- Prakken, H. (2020), An argumentation-based analysis of the Simonshaven case. *Topics in Cognitive Science* 12 (4), 1068-1091. (Special issue on Models of Rational Proof in Criminal Law.)
- Prakken, H. (2021), Philosophical reflections on argument strength and gradual acceptability. In J. Vejnárova & N. Wilson (eds.): *Proceedings of the 16th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 144-158. Springer Lecture Notes in AI 12897, Springer Verlag, Berlin.
- Prakken, H. (2026), Recognising and avoiding fallacies in interpreting statistical evidence. To appear in S. Brewer, J. Maranhao & G. Sartor (eds), *Logic and Law*. Edward Elgar Publishing.
- Timmer, S.T., Meyer, J.-J.Ch., Prakken, H., Renooij, S. & Verheij, B. (2017), A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning* 80, 475-494.
- Toulmin, S.E. (1958), *The Uses of Argument*. Cambridge University Press.
- Walton, D.N., (1996), *Argumentation Schemes for Presumptive Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zenker, F., Debowska-Kozłowska, K., Godden, D., Selinger & M., Wells, S. (2020), Five approaches to argument strength: probabilistic, dialectical, structural, empirical, and computational. In: *Proceedings of the 3rd European Conference on Argumentation*. pp. 653–674. College Publications, London.
- Zenker, F. (ed.) (2013), *Bayesian Argumentation. The Practical Side of Probability*. Springer, Dordrecht.