International Journal of Approximate Reasoning 146 (2022) 79–105

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Argument strength in probabilistic argumentation based on defeasible rules

Anthony Hunter

Department of Computer Science, University College London, UK

ARTICLE INFO

Article history:
Received 9 December 2021
Received in revised form 11 March 2022
Accepted 7 April 2022
Available online 11 April 2022

Keywords:
Computational argumentation
Probabilistic argumentation
Argument strength

ABSTRACT

It is common for people to remark that a particular argument is a strong (or weak) argument. Having a handle on the relative strengths of arguments can help in deciding on which arguments to consider, which arguments to regard as acceptable, and on which arguments to present to others in a discussion. In computational models of argument, there is a need for a deeper understanding of argument strength. It is a multidimensional problem, and in this paper, we focus on one aspect of argument strength for deductive argumentation based on a defeasible logic. We assume a probability distribution over models of the language and consider how there are various ways to calculate argument strength based on the probabilistic necessity and sufficiency of the premises for the claim, the probabilistic sufficiency of competing premises the claim, and the probabilistic necessity of the premises for competing claims. We provide axioms for characterizing probability-based measures of argument strength, and we investigate four specific probability-based measures.

1. Introduction

In real-world argumentation, it is common for arguments to be considered in terms of their strength. Yet in computational models of argument, we lack formalisms that adequately measure strength in arguments.

1.1. Motivation

For computational models of argument, having a better understanding of argument strength would help us to identify the arguments that we need to focus on when making a decision, and help us to be more effective in dialogues where we may be aiming to win a discussion or persuade someone to accept an argument. Having an intended recipient of an argument acknowledge that it is a strong argument can be useful in winning in a dialogue, whereas having them regard it as a weak argument can have the opposite effect. Also arguing about whether an argument is strong or weak leads to more sophisticated formalisms for computational argumentation, e.g. [12].

Argument strength is a complex concept that combines features of arguments and their context including: the **premises**, i.e. the quality of the premises (such as the reliability of the evidence or knowledge used); the **derivation**, i.e. the quality of the derivation of the claim (which may be defeasible or uncertain); the **dialectical criteria**, i.e. the quality of counterarguments and supporting arguments that build on notions of acceptability (see [20,15,5] for some key approaches to

E-mail address: anthony.hunter@ucl.ac.uk.

https://doi.org/10.1016/j.ijar.2022.04.003

acceptability criteria); the **interaction context**, i.e. the broader context of the argumentation (such as what has been previously said about the topic of the argumentation and by whom, what concessions have been made, etc); and the **audience context**, i.e. the perception of the arguments from the perspective of the intended audience (such as what the audience already believes, what concerns they have (e.g. [2,26]), and what kinds of emotional response they might have to the arguments [38,29,4]). For reviews of notions of argument strength in the literature on computational models of argument, see [53,8].

Our focus in this paper is on the premises and derivation of arguments. In most abstract and logic-based (i.e. structured) approaches to argumentation, the assessment of these dimensions is pushed to the dialectical level. In other words, if there is doubt about the strength of an argument because of the quality of premises, claim, or derivation, then this is expressed via counterarguments. But this does not involve a general understanding of what a strong argument is. It assumes somehow that the counterarguments can be found.

Some variants of abstract argumentation touch on the notion of strength such as rankings (e.g. [15,42,1,3,10]), and probabilities (e.g. [19,39,60,33,55,32]). However, these do not capture a notion of argument strength in terms of the quality of the contents of the premises and/or claim, rather they either assume that some kind of strength value is given for each argument and/or they calculate strength in terms of attacking and supporting arguments.

Using logical (i.e. structured) arguments allows the quality of the contents of the premises and claim to be directly considered. Some proposals assume strength is an input to the system (e.g. [12]). Others assess the strength of an argument in terms of the belief in it, often in terms of belief in the premises and claim (e.g. [25,33,52]), or in terms of the conditional probability of the claim given the premises of the argument (e.g. [50,62,34,18]). So these draw on uncertainty in argumentation to quantify the strength.

But taking the belief of the premises into account fails to consider the conditionality of the argument. For example, consider the argument "*it is raining now in the Sahara, so the ground there will be wet*". The probability of the premise (i.e. "*it is raining now in the Sahara*") being true is low, but if the premise was true, then the probability of the claim (i.e. "*the ground there will be wet*") being true would be high. This can be measured by considering the probability of the claim given the premise, i.e. as a conditional probability.

Looking at the conditional probability of the premise ψ given the claim ϕ can be illuminating (i.e. $P(\phi|\psi)$). However, the conditional probability can be misleading and we may want to also look at the sufficiency of the complementary premise for the claim (i.e. $P(\phi|\neg\psi)$). We may also want to consider the necessity of the premise for the claim (i.e. $P(\psi|\phi)$) and the necessity of the premises for the complementary claim (i.e. $P(\psi|\neg\phi)$). We draw on these dimensions in the following examples, and explore the dimensions systematically in the rest of the paper.

- **Sufficiency of premise for the claim.** For instance, the premise "*she gets good grades in her coursework*" would appear to score highly for sufficiency for the claim "*she will get good grades in her exams*" (i.e. the probability of the latter conditioned on the former), whereas the premise "*she plays a lot of football*" would not appear to score highly for sufficiency for the claim "*she will get good grades in her exams*". So for a typical probability distribution, we would expect that the conditional probability would show that indeed "*she gets good grades in her coursework*" is highly sufficient for the claim "*she will get good grades in her exams*" whereas "*she plays a lot of football*" would not.
- **Sufficiency of complementary premise for the claim.** For instance, the conditional probability for the claim "*the tide comes in*" given the evidence "*it is sunny*" is 1, and so suggests the premise is completely sufficient for the claim, but the conditional probability for the claim "*the tide comes in*" given the evidence "*it is not sunny*" is also 1 (i.e. the conditional probability of the complementary premise for the claim is 1). So the claim "*the tide comes in*" holds irrespective of whether or not "*it is sunny*" holds.
- **Necessity of premise for the claim.** For instance, the premise "*strong winds from the south*" might score highly for being sufficient for the claim "*it will rain soon*", but it might be an uncommon meteorological scenario and so might not have a high degree of necessity for the claim. In other words, the premise conditioned on the claim might be low. Now consider the premise "*dark clouds*" which might occur relatively often, and when it does, the claim "*it will rain soon*" normally holds. So "*dark clouds*" scores highly for being sufficient for the claim. Furthermore, if most of the time that "*it will rain soon*" holds, "*dark clouds*" holds, then "*dark clouds*" also scores highly for being necessary for the claim. We can then say that in this probabilistic sense, "*dark clouds*" is a better explanation than "*strong winds from the south*" for the claim "*it will rain soon*" holding.
- **Necessity of premise for the complementary claim.** For instance, consider a scenario of an unusual zoo with just three kinds of animal (namely, bats, hawks, and penguins). Furthermore, suppose that there are 900 bats, 90 hawks, and 10 penguins. Let "*bird*" be the premise and let "*capable of flying*" be the claim. So the premise has a high degree of sufficiency for the claim (i.e. most of the birds in the zoo are capable of flying), but the degree of necessity of the premise for the claim is low (i.e. most animals in the zoo that are capable of flying are not birds), whereas the degree of necessity of the premise for the negation of the claim is high (i.e. all the animals that are not capable of flying are birds). So in this probabilistic sense, "*bird*" is not a good explanation for "*capable of flying*" but "*bird*" is a good explanation for the negation of "*capable of flying*".

The above examples illustrate how different kinds of conditional probability statement can provide valuable insights into the connection between premises and claims in argumentation. This then raises interesting questions about how we can harness these in structured argumentation which we summarize as goals in the next section.

1.2. Overview of proposal

Following the motivation in the previous section, we will investigate how we can use probabilistic information to systematically investigate the connection between the premises and claim of a logical argument with the aim of providing probability-based measures of argument strength.

1.2.1. Key goals

In this paper, we will have the following sequence of goals so that the solutions to the early goals will be used in addressing the later goals. In the body of the paper, we will refer to these goals in order to navigate the presentation.

Goal 1: Probabilistic argumentation based on defeasible logic We want a probabilistic defeasible logic with proof theory and semantics that can be used as the base logic for instantiating probabilistic argumentation. The logic needs to support a clear representation of uncertainty so that it is straightforward to analyze uncertainty in structured argumentation. We can then use this probabilistic argumentation to investigate notions of argument strength in the following objectives.

Goal 2: Framework of relationships between evidence and claims We want to draw out four dimensions of an argument, as indicated in the previous section: (1) the sufficiency of the premises for the claim; (2) the necessity of the premises for the claim; (3) the sufficiency of competing premises for the claim; and (4) the necessity of the premises for a competing claim. Looking at all these dimensions provides us with a more complete picture of the connection between belief in the premises and belief in the claim of an argument.

Goal 3: Framework of axioms for argument strength We want general definitions of a measure of argument strength, and of a probability-based measure of argument strength, and then we want to consider a framework of axioms that can constrain, or characterize, specific measures of argument strength, and in particular, probability-based measures of argument strength.

Goal 4: Proposals for measures of argument strength We want to define some probability-based measures for probabilistic argumentation based on defeasible logic (i.e. the outcome of addressing objective 1). We want the measures to be defined in terms of the framework of relationships between evidence and claims (i.e. the outcome of addressing objective 2), and we want to investigate the proposals using the framework of axioms for argument strength (i.e. the outcome of addressing objective 3).

The proposal in this paper (i.e. the solutions to the above goals) could be used in various ways. To illustrate and motivate, we can consider the following **audience scenario**: Someone presents us with an argument graph and the knowledgebase from which it has been constructed. The knowledge represents patterns that normally hold in the world (e.g. "*if it is bird, then it is capable of flying*"). As a member of the audience, we are at liberty to identify a probability distribution over the possible worlds (in the following sections we make this precise) that represent our beliefs about the propositions in the language, and then we can use this probability distribution to analyze the strength of the arguments presented to us.

Other applications include a **persuasion scenario** (for a review of persuasion see [26]) where we can construct a probability distribution that reflects what we think the other agent believes about the world, and then use this to select the stronger arguments to present, and an **analytical scenario** where a probability distribution could reflect a possible modelling of the world, and so the change in strength for specific arguments could be investigated for different probability distributions. We will return to these scenarios in the discussion.

1.2.2. Background

In order to address the above goals, we will we use a variant of defeasible logic as the base logic (where a base logic is the logic used to generate arguments and counterarguments [7]). The use of defeasible logic is well-established in argumentation (see for example [49,57,22]), and key approaches to structured argumentation incorporate various kinds of defeasible rules [6]. There are also some proposals for probabilistic quantification of uncertainty in argumentation systems based on defeasible logic (e.g. [56,19,58]), but quantifying notions of strength have not been systematically considered in these formalisms. In order to introduce a probabilistic version of defeasible logic, we adapt the epistemic approach to probabilistic argumentation [60,33,32] and this will enable us to address Goal 1.

Based on the epistemic approach to probabilistic argumentation, we can investigate how we can use conditional probabilities to investigate the connection between the premises and claim (i.e. Goal 2). The need to consider the dimensions outlined in Goal 2 has been influenced by research in confirmation theory. Originally, confirmation measures were developed in the philosophy of science to investigate the development of scientific hypotheses [17]. The aim of a confirmation measure $C(E, H)$ is to capture the degree to which evidence E supports hypothesis H. Some proposals for confirmation are the following: C^d_P [11], C^s_P [14], C^n_P [44], and C^k_P [37], where $\overline{H}$ is the negation of H and $\overline{E}$ is the negation of E.

- $C^d_P(E,H) = P(H|E) - P(E)$
- $C^s_P(E,H) = P(H|E) - P(H|\overline{E})$
- $C^n_P(E,H) = P(E|H) - P(E|\overline{H})$
- $C^k_P(E,H) = \frac{P(E|H)-P(E|\overline{H})}{P(E|H)+P(E|\overline{H})}$

We explain these measures as follows: (C^d_P) the increase in belief in the hypothesis that can be attributed to believing the evidence to be true, (C^s_P) the difference in belief in the hypothesis conditioned on the evidence being true and belief in the hypothesis conditioned on the evidence being untrue, (C^n_P) the difference in belief in the evidence conditioned on the hypothesis being true and belief in the evidence conditioned on the hypothesis being untrue, and (C^k_P) the difference in belief in the hypothesis conditioned on the evidence being true and belief in the hypothesis conditioned on the evidence being untrue, normalized by the maximum range for the value.

Whilst confirmation measures have been suggested as a measure of the strength of arguments [46], they have not been systematically considered for formalisms for computational argumentation. In particular, there is a need to consider how they can be applied to structured argumentation, and to do this systematically, it is better to start with a consideration of the different ways that conditional probabilities can be used to capture the relationships between the premises and claim of an argument (i.e. Goal 2), and to then propose axioms to constrain and characterize measures (i.e. Goal 3), before considering specific measures (i.e. Goal 4).

One of the underlying reasons that we consider necessity of premises for a claim, and necessity of premises for the complementary claim, is that we are interested in whether a premise is a good explanation for a claim. This is a specific sense of the notion of explanation, and we will investigate it later in the paper. Whilst this notion of explanation is related to broader efforts for explanations in argumentation (see [16] for a review), we do not consider these broader issues in this paper.

1.2.3. Overview of rest of paper

We proceed as follows: (Section 2) We review some preliminaries on probabilistic qualification of logical reasoning; (Section 3) We introduce the proof theory and semantics for a variant of defeasible logic (*Goal 1*); (Section 4) We use the defeasible logic from the previous section as the base logic for probabilistic argumentation (*Goal 1*); (Section 5) We consider modelling of normality in probabilistic argumentation (*Goal 1*); (Section 6) We investigate conditional probability statements as a way of capturing sufficiency and necessity in reasoning (*Goal 2*); (Section 7) We present a set of properties that characterize aspects of probability-based measures of argument strength (*Goal 3*); (Section 8) We present four probability-based measures of argument strength (*Goal 4*); (Section 9) We investigate the measures presented in Section 8 with respect to the properties presented in Section 7 (*Goal 4*); (Section 10) We consider how we can analyze arguments with multiple defeasible rules (*Goal 1*); (Section 11) We consider how we can update probability distributions (*Goal 1*); And (Section 12) we conclude and consider future work.

Note, this paper is a development of the research presented in [35]. The same defeasible logic is used. However, in this paper, we provide the following novel contributions: A systematic analysis of how conditional probabilities can capture different aspects of the sufficiency and necessity of premises for claims; A set of properties that characterize different kinds of strength measures; And a set of four strength measures that are defined in a systematic way based on conditional probabilities and that we analyze using the characterization properties. In contrast, in [35], the strength measures were directly taken from the confirmation theory literature, and so it was difficult to compare different measures in the context of argumentation.

2. Preliminaries

In these preliminaries, we will review how we can augment classical propositional logic with probabilities. So the language is classical propositional logic, and the probability distribution is over two-valued models. This is a well-known approach to capturing probabilities in propositional logic [9,45,43].

We assume a finite **set of atoms** $\mathcal{A}$. The set of models based on $\mathcal{A}$ is denoted $\mathcal{M}(\mathcal{A})$ and is defined a follows: We use a **signature**, denoted $\mathcal{S}$, which is the set of atoms $\mathcal{A}$ given in a sequence $\langle a_1, ..., a_n \rangle$, and then each model $m \in \mathcal{M}(\mathcal{A})$ is a binary number $b_1...b_n$ where for each digit b_i, if b_i is 1, then a_i is true in m, and if b_i is 0, then a_i is false in m. So if the cardinality of $\mathcal{A}$ is n, then $\mathcal{M}(\mathcal{A})$ is the set of n-digit binary numbers.

Example 1. For $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, where $\mathcal{S} = \langle \mathtt{a}, \mathtt{b}, \mathtt{c} \rangle$, $\mathcal{M}(\mathcal{A})$ is $\{111, 110, 101, 100, 011, 010, 001, 000\}$. So for $m = 101$, $\mathtt{a}$ is true, $\mathtt{b}$ is false, and $\mathtt{c}$ is true.

A probability distribution over the models is an assignment of a value in the unit interval to each model such that the sum of the assignments is 1.

Definition 1. A **probability distribution** P over $\mathcal{M}(\mathcal{A})$ is a function $P : \mathcal{M}(\mathcal{A}) \rightarrow [0, 1]$ s.t. $\sum_{m \in \mathcal{M}(\mathcal{A})} P(m) = 1$.

Example 2. Continuing Example 1, the following is a probability distribution: $P(111) = 0.5$, $P(110) = 0.1$, $P(010) = 0.2$, and $P(001) = 0.2$.

To simplify some definitions in this paper, we also require the language of classical propositional logic. For a set of atoms $\mathcal{A}$, let $\mathcal{B}(\mathcal{A})$ be the set of classical propositional formulae that can be formed from the atoms using the usual negation, disjunction, and conjunction, connectives (i.e. $\mathcal{A} \subseteq \mathcal{B}(\mathcal{A})$ and for $\phi, \psi \in \mathcal{B}(\mathcal{A})$, $\neg\phi, \phi \vee \psi, \phi \wedge \psi \in \mathcal{B}(\mathcal{A})$). We also require the notion of classical satisfiability for these formulae. For a formula $\phi \in \mathcal{B}(\mathcal{A})$, let $\mathsf{Sat}(\phi)$ be the set of models that satisfy ϕ and this is defined in the usual way for classical propositional logic: For an atom $\phi \in \mathcal{B}(\mathcal{A})$, $\mathsf{Sat}(\phi)$ is the set of models such that if ϕ is the ith atom in $\mathcal{S}$, then each model in $\mathsf{Sat}(\phi)$ has 1 for the ith digit (i.e. $\mathsf{Sat}(\phi) = \{m \in \mathcal{M}(\mathcal{A}) \mid a_i \in \langle a_1, \ldots, a_n \rangle$ and $\phi = a_i$ and $m = b_1 \ldots b_n$ and $b_i = 1\}$), and then we treat arbitrary propositional formulae by recursion: $\mathsf{Sat}(\phi \wedge \psi) = \mathsf{Sat}(\phi) \cap \mathsf{Sat}(\psi)$, $\mathsf{Sat}(\phi \vee \psi) = \mathsf{Sat}(\phi) \cup \mathsf{Sat}(\psi)$, and $\mathsf{Sat}(\neg\phi) = \mathcal{M}(\mathcal{A}) \setminus \mathsf{Sat}(\phi)$. Also, for $\phi, \psi \in \mathcal{B}(\mathcal{A})$, let $\phi \equiv \psi$ denote that ϕ and ψ are equivalent (i.e. $\mathsf{Sat}(\phi) = \mathsf{Sat}(\psi)$).

Example 3. Continuing Example 1, $\mathsf{Sat}(\mathtt{a}) = \{111, 110, 101, 100\}$, $\mathsf{Sat}(\neg\mathtt{a}) = \{011, 010, 001, 000\}$, $\mathsf{Sat}((\mathtt{b} \vee \mathtt{c})) = \{111, 110, 101, 011, 010, 001\}$, and $\mathsf{Sat}(\neg\mathtt{a} \wedge (\mathtt{b} \vee \mathtt{c})) = \{011, 010, 001\}$.

The probability of a formula is the sum of the probability assigned to the models that satisfy the formula.

Definition 2. The **probability of a formula** $\phi \in \mathcal{B}(\mathcal{A})$ w.r.t. a probability distribution P is $P(\phi) = \sum_{m \in \mathsf{Sat}(\phi)} P(m)$.

Example 4. Continuing Example 2 and Example 3, $P(\neg\mathtt{a} \wedge (\mathtt{b} \vee \mathtt{c})) = P(010) + P(001) = 0.4$.

A **conditional probability**, denoted $P(\phi|\psi)$, is $P(\phi \wedge \psi)/P(\psi)$ if $P(\psi) > 0$, and $P(\phi|\psi)$ is undefined otherwise.

Example 5. Continuing Example 2, $P(\mathtt{a}|\mathtt{b} \vee \mathtt{c}) = 0.6$

We conclude this section with some lemmas concerning conditional probabilities that we will use in subsequent sections.

Lemma 1. *For a probability distribution P, and formulae $\phi, \psi \in \mathcal{B}(\mathcal{A})$. $P(\phi|\neg\psi) \# P(\phi|\psi)$ iff $P(\phi) \# P(\phi|\psi)$, where $\# \in \{<, \leq, =, \geq, >\}$.*

Lemma 2. *For a probability distribution P, and formulae $\phi, \psi \in \mathcal{B}(\mathcal{A})$. $P(\phi|\psi) \# P(\phi)$ iff $P(\psi|\phi) \# P(\psi)$, where $\# \in \{<, \leq, =, \geq, >\}$.*

Lemma 3. *For a probability distribution P, and formulae $\phi, \psi \in \mathcal{B}(\mathcal{A})$. $P(\phi|\psi) \# P(\phi|\neg\psi)$ iff $P(\psi|\phi) \# P(\psi|\neg\phi)$ where $\# \in \{<, \leq, =, \geq, >\}$.*

Note, the proofs for these lemmas, and for all propositions in the rest of the paper, can be found in the Appendix of the paper.

3. Defeasible logic

In this section, we present a variant of defeasible logic that we will use as a base logic for probabilistic argumentation. This is the first part of addressing Goal 1.

The language is a sublanguage, and the proof theory is a subsystem, of classical propositional logic. The semantics diverges from classical propositional logic but we will use the two-valued models reviewed in the previous section for the semantics for the defeasible logic (as explained below).

We start with the language for our variant of propositional defeasible logic. From a finite set of atoms $\mathcal{A}$, we form a **set of literals** $\mathcal{L}(\mathcal{A}) = \mathcal{A} \cup \{\neg\phi \mid \phi \in \mathcal{A}\}$. A **defeasible rule** is of the form $\psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi$ where $\psi_1, \ldots, \psi_n, \phi \in \mathcal{L}(\mathcal{A})$. For a rule ρ of the form $\psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi$, let $\mathsf{Tail}(\rho) = \{\psi_1, \ldots, \psi_n\}$ and $\mathsf{Head}(\rho) = \phi$. The **set of rules** is $\mathcal{R}(\mathcal{A})$ and the **set of formulae** is $\mathcal{F}(\mathcal{A}) = \mathcal{L}(\mathcal{A}) \cup \mathcal{R}(\mathcal{A})$. A **knowledgebase** is a subset of $\mathcal{F}(\mathcal{A})$.

Example 6. Consider $\mathtt{b}$ for *bird*, $\mathtt{p}$ for *penguin*, and $\mathtt{f}$ for *capable of flying*. Then $\Delta = \{\mathtt{b}, \mathtt{p}, \mathtt{b} \rightarrow \mathtt{f}, \mathtt{p} \rightarrow \neg\mathtt{f}\} \subseteq \mathcal{F}(\mathcal{A})$ is a knowledgebase.

Next, we present the consequence relation for a variant of defeasible logic, incorporating the *modus ponens* proof rule (rule 2 below) and *ex falso quodlibet* proof rule (rule 3 below).

Definition 3. Let $\Delta \subseteq \mathcal{F}(\mathcal{A})$ be a knowledgebase and $\phi, \psi \in \mathcal{L}(\mathcal{A})$. The **consequence relation**, denoted $\vdash$, is defined as follows: (1) $\Delta \vdash \phi$ if $\phi \in \Delta$; (2) $\Delta \vdash \phi$ if there is a $\psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi \in \Delta$ and $\Delta \vdash \psi_1$ and $\ldots$ and $\Delta \vdash \psi_n$; and (3) $\Delta \vdash \phi$ if $\Delta \vdash \psi$ and $\Delta \vdash \neg\psi$. Let $\mathsf{Closure}(\Delta) = \{\phi \in \mathcal{A} \mid \Delta \vdash \phi\}$.

Example 7. For $\Delta = \{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{f}\}$, $\mathsf{Closure}(\Delta) = \{\mathtt{b}, \mathtt{f}\}$.

The Tarskian properties (widely regarded as requirements for a logic) are satisfied (though for reflexivity, it is restricted to literals): (Reflexivity) $(\Delta \cap \mathcal{L}(\mathcal{A})) \subseteq \mathsf{Closure}(\Delta)$; (Monotonicity) $\mathsf{Closure}(\Delta) \subseteq \mathsf{Closure}(\Delta')$ if $\Delta \subseteq \Delta'$; and (Idempotency) $\mathsf{Closure}(\Delta) \subseteq \mathsf{Closure}(\mathsf{Closure}(\Delta))$.

For a set of atoms $\mathcal{A}$, $\mathcal{M}(\mathcal{A})$ is the set of models as defined in the previous section. However, the notion of satisfaction for the language of defeasible logic $\mathcal{F}(\mathcal{A})$ is different to that given for classical propositional logic $\mathcal{B}(\mathcal{A})$ given in the previous section. To define satisfaction, we use the following fixpoint of the Inf function where the Inf function generates literals by the application of defeasible rules.

Definition 4. For $\Delta \subseteq \mathcal{F}(\mathcal{A})$, and $i \in \mathbb{N}$, the **inference operators**, denoted Inf^i, is defined as: $\mathsf{Inf}^1(\Delta) = \Delta \cap \mathcal{L}(\mathcal{A})$ and $\mathsf{Inf}^{i+1}(\Delta) = \mathsf{Inf}^i(\Delta) \cup \{\mathsf{Head}(\rho) \mid \rho \in \Delta \cap \mathcal{R}(\mathcal{A})$ and for all $\psi \in \mathsf{Tail}(\rho), \psi \in \mathsf{Inf}^i(\Delta)\}$. Let $\mathsf{Infer}(\Delta) = \mathsf{Inf}^k(\Delta)$ where k is the smallest value s.t. $\mathsf{Inf}^k(\Delta) = \mathsf{Inf}^{k+1}(\Delta)$.

The satisfying models for a knowledgebase $\Delta \subseteq \mathcal{F}(\mathcal{A})$, denoted $\mathsf{Models}(\Delta)$, is the set of models that satisfy all the literals in $\mathsf{Infer}(\Delta)$. So satisfaction for the defeasible logic is defined in terms of the satisfaction of literals for classical logic as follows.

Definition 5. The **satisfying models** for $\Delta \subseteq \mathcal{F}(\mathcal{A})$, is $\mathsf{Models}(\Delta) = \bigcap_{\phi \in \mathsf{Infer}(\Delta)} \mathsf{Sat}(\phi)$.

Example 8. For $\mathcal{A} = \{\mathtt{a}\}$, let $\Delta_1 = \{\mathtt{a} \rightarrow \mathtt{a}\}$, $\Delta_2 = \{\mathtt{a}, \mathtt{a} \rightarrow \mathtt{a}\}$, $\Delta_3 = \{\mathtt{a}, \mathtt{a} \rightarrow \neg\mathtt{a}\}$, $\Delta_4 = \{\neg\mathtt{a}, \neg\mathtt{a} \rightarrow \mathtt{a}\}$, and $\Delta_5 = \{\neg\mathtt{a}\}$.

	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
$\mathsf{Infer}(\Delta_i)$	$\emptyset$	$\{\mathtt{a}\}$	$\{\mathtt{a}, \neg\mathtt{a}\}$	$\{\mathtt{a}, \neg\mathtt{a}\}$	$\{\neg\mathtt{a}\}$
$\mathsf{Models}(\Delta_i)$	$\{1, 0\}$	$\{1\}$	$\emptyset$	$\emptyset$	$\{0\}$

Example 9. For $\Delta = \{\mathtt{b}, \neg\mathtt{o}, \mathtt{b} \rightarrow \mathtt{f}, \mathtt{p} \rightarrow \neg\mathtt{f}\}$, where $\mathcal{A} = \{\mathtt{b}, \mathtt{p}, \mathtt{f}, \mathtt{o}\}$, $\mathcal{S} = \langle \mathtt{b}, \mathtt{p}, \mathtt{f}, \mathtt{o} \rangle$, $\mathsf{Infer}(\Delta) = \{\mathtt{b}, \neg\mathtt{o}, \mathtt{f}\}$, and $\mathsf{Models}(\Delta) = \{1010, 1110\}$.

It is straightforward to show that for a defeasible rule $\rho \in \mathcal{R}(\mathcal{A})$, if $\mathsf{Tail}(\rho) \subseteq \mathsf{Infer}(\Delta)$, then $\mathsf{Models}(\Delta \cup \{\rho\}) = \mathsf{Models}(\Delta \cup \{\mathsf{Head}(\rho)\})$, whereas if $\mathsf{Tail}(\rho) \nsubseteq \mathsf{Infer}(\Delta)$, then $\mathsf{Models}(\Delta \cup \{\rho\}) = \mathsf{Models}(\Delta)$.

Definition 6. For $\Delta \subseteq \mathcal{F}(\mathcal{A})$, $\phi \in \mathcal{L}(\mathcal{A})$, the **entailment relation** holds, denoted $\Delta \models \phi$, iff $\mathsf{Models}(\Delta) \subseteq \mathsf{Models}(\phi)$.

Example 10. For knowledgebase $\Delta = \{\mathtt{b}, \mathtt{p}, \mathtt{b} \rightarrow \mathtt{f}\}$, $\Delta \models \mathtt{f}$, but for $\Delta = \{\mathtt{f} \rightarrow \mathtt{f}\}$, $\Delta \not\models \mathtt{f}$.

A knowledgebase is **consistent** iff it does not imply an atom and its negation. So Example 6 (respectively Example 10) is inconsistent (respectively consistent). Obviously, Δ is consistent iff $\mathsf{Models}(\Delta) \neq \emptyset$.

The following result shows that the proof theory is sound and complete.

Proposition 1. *For $\Delta \subseteq \mathcal{F}(\mathcal{A})$, $\phi \in \mathcal{L}(\mathcal{A})$, $\Delta \vdash \phi$ iff $\Delta \models \phi$.*

Our defeasible logic is similar to previous proposals for defeasible logic in that it is based on a language of literals and rules, and the main proof rule is *modus ponens*. However, we also use *ex falso quodlibet* as a proof rule. This allows to also have a simple semantics which will be useful when we consider using the logic as a base logic for probabilistic argumentation as discussed in the next section.

4. Probabilistic argumentation

In this section, we adapt deductive argumentation (for a review see [7]), and the epistemic approach to probabilistic argumentation [60,33,32], for use with defeasible logic. This will address the second part of Goal 1.

An argument is a set of premises from $\mathcal{F}(\mathcal{A})$, and a claim from $\mathcal{L}(\mathcal{A})$, such that (1) the premises imply the claim; (2) the premises are consistent; and (3) the premises are minimal for entailing the claim.

Definition 7. For $\Phi \subseteq \mathcal{F}(\mathcal{A})$, and $\alpha \in \mathcal{L}(\mathcal{A})$, $\langle \Phi, \alpha \rangle$ is an **argument** iff (1) $\Phi \vdash \alpha$; (2) Φ is consistent; and (3) there is no $\Phi' \subset \Phi$ such that $\Phi' \vdash \alpha$.

Example 11. For $\Delta = \{\mathtt{b}, \mathtt{p}, \mathtt{b} \rightarrow \mathtt{f}, \mathtt{p} \rightarrow \neg\mathtt{f}\}$, the arguments are $\langle \{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{f}\}, \mathtt{f} \rangle$, $\langle \{\mathtt{p}, \mathtt{p} \rightarrow \neg\mathtt{f}\}, \neg\mathtt{f} \rangle$, $\langle \{\mathtt{b}\}, \mathtt{b} \rangle$, and $\langle \{\mathtt{p}\}, \mathtt{p} \rangle$.

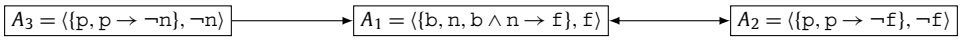

Fig. 1. An argument graph from $\Delta = \{\mathtt{b}, \mathtt{p}, \mathtt{b} \wedge \mathtt{n} \rightarrow \mathtt{f}, \mathtt{p} \rightarrow \neg\mathtt{f}, \mathtt{p} \rightarrow \neg\mathtt{n}\}$ where $\mathtt{b}$ is *bird*, $\mathtt{p}$ is *penguin*, $\mathtt{f}$ is *capable of flying*, and $\mathtt{n}$ is a normality atom. Argument A_3 has the claim that negates $\mathtt{n}$, and so represents an attack on the use of the rule with $\mathtt{n}$ as condition in A_1.

For argument $A = \langle\Phi, \alpha\rangle$, Φ is the **support**, and α is the **claim**, of the argument. Also for argument $A = \langle\Phi, \alpha\rangle$, $\mathsf{Support}(A)$ returns Φ, $\mathsf{C}(A)$ returns α, $\mathsf{Facts}(A)$ returns the literals in Φ, $\mathsf{Rules}(A)$ returns the defeasible rules in Φ, and $\mathsf{Frame}(A)$ returns $\mathsf{Facts}(A) \cup \{\mathsf{C}(A)\}$. So the frame of the argument is all the literals in the premises and claim of the argument. Note, we have a shorter name for C because we will be using it many times through the paper.

The probability of an argument being acceptable is based on the facts in the support and the claim of the argument. So returning to the audience scenario we considered in the introduction, if the audience has been presented with an argument graph, they can use their probability distribution over the models to calculate the probability of each argument.

Definition 8. Let P be a probability distribution over $\mathcal{M}(\mathcal{A})$. The **probability of argument** A being acceptable is denoted $P(A)$, where $P(A) = \sum_{m \in \mathsf{Models}(\mathsf{Frame}(A))} P(m)$.

Example 12. Continuing Example 6, with probability distribution P below, and signature $\mathcal{S} = \langle\mathtt{b}, \mathtt{p}, \mathtt{f}\rangle$, $P(\langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle) = 0.95$ and $P(\langle\{\mathtt{p}, \mathtt{p} \rightarrow \neg\mathtt{f}\}, \neg\mathtt{f}\rangle) = 0.01$.

$\langle\mathtt{b}, \mathtt{p}, \mathtt{f}\rangle$	110	101	100
P	0.01	0.95	0.04

Next we consider some notions of counterargument that are adapted from deductive argumentation [7].

Definition 9. For $\phi \in \mathcal{L}(\mathcal{A})$, let $\neg\neg\phi = \phi$ (so we can rewrite $\neg\neg\phi$ to ϕ). For arguments A and B, A is a **direct undercut** of B if there is $\phi \in \mathsf{Support}(B)$ s.t. $\mathsf{C}(A) \equiv \neg\phi$. A is a **rebuttal** of B if $\mathsf{C}(A) \equiv \neg\mathsf{C}(B)$. A **attacks** B iff A is a direct undercut or A is a rebuttal of B.

The following coherence property holds because the sum of belief in complementary literals is less than or equal to 1.

Proposition 2. *For probability distribution P, if B attacks A, then $P(A) + P(B) \leq 1$.*

We use the usual notion of an argument graph, where each node is an argument, and each arc denotes an attack by one argument on another [20] (e.g. Fig. 1). However, we are not using the argument graph to determine which arguments are acceptable. Rather, we use a probability distribution to determine the acceptable arguments as explained below. So the role of the argument graph is to provide a presentation of the arguments.

When $P(A) > 0.5$, then the argument is believed to be acceptable, whereas when $P(A) \leq 0.5$, then it is not believed to be acceptable. The epistemic extension for a graph G, denoted $\mathsf{Extension}(P, G)$, is the set of arguments that are believed to be acceptable (i.e. $A \in \mathsf{Extension}(P, G)$ iff A is in G and $P(A) > 0.5$).

Example 13. For graph G in Fig. 1, with $P(A_1) = 0.95$, and $P(A_2) = P(A_3) = 0.15$, $\mathsf{Extension}(P, G) = \{A_1\}$.

As shown in [33], for any probability distribution P and graph G, $\mathsf{Extension}(P, G)$ is conflict-free. However, the epistemic approach provides a finer-grained assessment of an argument graph than given by the definitions for Dung's extensions. By adopting constraints on the distribution, the epistemic approach subsumes Dung's definitions [32,48]. Importantly, the epistemic approach also provides alternatives to Dung's approach. For instance, we may wish to represent disbelief in arguments even when they are unattacked [47]. This is particularly valuable when we are unable to add extra arguments to the argument graph, and so we are only able to express belief in those arguments that are in the argument graph.

5. Modelling normality

Defeasible rules are normally correct, but sometimes are incorrect, and so we may need to attack them. To address this, we can use normality atoms. We assume the set of atoms $\mathcal{A}$ is partitioned into **normality atoms**, denoted $\mathcal{N}$, and **ordinary atoms**, denoted $\mathcal{Q}$. So $\mathcal{A} = \mathcal{Q} \cup \mathcal{N}$ and $\mathcal{Q} \cap \mathcal{N} = \emptyset$. We read the normality atom in the condition of a defeasible rule as saying that the context for applying the defeasible rule is normal. If there are reasons to believe that it is not a normal context, then a counterargument attacks this assumption of normality. We illustrate the use of normality atoms in the following example and in Fig. 1.

Example 14. Let $\mathcal{Q} = \{\mathtt{wd}, \mathtt{wd}, \mathtt{um}\}$ and $\mathcal{N} = \{\mathtt{nd}\}$. Argument A_1 captures the general rule that if $\mathtt{wd}$ (i.e. *it is a workday*) holds, then $\mathtt{um}$ (i.e. *use metro*) holds. The use of the rule in A_1 requires that the assumption $\mathtt{nd}$ (i.e. *it is a normal day*) holds.

This is given as a premise. Argument A_2 undercuts the first argument by contradicting the assumption that nd holds based on the premise wh (i.e. *it is a work at home day*).

$$A_1 = \langle \{\mathtt{wd}, \mathtt{nd}, \mathtt{wd} \wedge \mathtt{nd} \rightarrow \mathtt{um}\}, \mathtt{um}) \rangle$$
$$A_2 = \langle \{\mathtt{wh}, \mathtt{wh} \rightarrow \neg \mathtt{nd}\}, \neg \mathtt{nd} \rangle$$

By having appropriate conditions in the antecedent of a defeasible rule we can disable the rule by generating a counterargument that attacks it. This in effect attacks the usage of the rule. This way to disable rules is analogous to the use of abnormality predicates in formalisms such as circumscription [40]. The use of normality atoms can also be viewed as analogous to the use of assumptions in assumption-based argumentation [61]. Furthermore, we can use normality atoms to capture the specificity principle where a more specific rule is preferred over a less specific rule (as illustrated below).

Example 15. (Adapted from [21]) Let $\mathcal{Q} = \{\mathtt{na}, \mathtt{c}, \mathtt{m}, \mathtt{s}\}$, where na is *nautilus*, c is *cephalopod*, m is *mollusc*, s is *shellbearer*, and $\mathcal{N} = \{\mathtt{n}_1, \mathtt{n}_2\}$. A nautilus is a type of cephalopod (i.e. $\mathtt{na} \rightarrow \mathtt{c}$) and a cephalopod is a type of mollusc (i.e. $\mathtt{c} \rightarrow \mathtt{m}$). Molluscs are normally shellbearers (i.e. $\mathtt{m} \wedge \mathtt{n}_1 \rightarrow \mathtt{s}$), but cephalopods are an exception for molluscs (i.e. $\mathtt{c} \wedge \mathtt{n}_2 \rightarrow \neg \mathtt{n}_1$), and nautilus is an exception for cephalopods (i.e. $\mathtt{na} \rightarrow \neg \mathtt{n}_2$). This gives the following arguments concerning nautilus being a shellbearer.

$$\langle \{\mathtt{na}, \mathtt{n}_1, \mathtt{na} \rightarrow \mathtt{c}, \mathtt{c} \rightarrow \mathtt{m}, \mathtt{m} \wedge \mathtt{n}_1 \rightarrow \mathtt{s}\}, \mathtt{s} \rangle$$
$$\langle \{\mathtt{na}, \mathtt{n}_2, \mathtt{na} \rightarrow \mathtt{c}, \mathtt{c} \wedge \mathtt{n}_2 \rightarrow \neg \mathtt{n}_1\}, \neg \mathtt{n}_1 \rangle$$
$$\langle \{\mathtt{na}, \mathtt{na} \rightarrow \neg \mathtt{n}_2\}, \neg \mathtt{n}_2 \rangle$$

To help us to specify an appropriate probability distribution over a set of atoms that includes normality atoms, we will use the following normality modelling convention which comes in three parts (defined next).

Definition 10. The **normality modelling convention** holds for a knowledgebase Δ iff for all rules $\rho, \rho' \in \mathcal{R}(\mathcal{A}) \cap \Delta$,

1. if $\gamma \in \mathsf{Tail}(\rho) \cap \mathcal{N}$ and $\gamma' \in \mathsf{Tail}(\rho) \cap \mathcal{N}$, then $\gamma = \gamma'$.
2. if $\gamma \in \mathsf{Tail}(\rho) \cap \mathcal{N}$ and $\gamma' \in \mathsf{Tail}(\rho') \cap \mathcal{N}$, and $\rho \neq \rho'$, then $\gamma \neq \gamma'$.
3. if $\mathsf{Head}(\rho) = \neg\gamma$ and $\gamma \in \mathcal{N}$, then $\gamma \in \mathsf{Tail}(\rho')$ for some $\rho' \in \Delta$, and $\gamma \notin \mathsf{Tail}(\rho)$.

In the first part of the convention, each rule has at most one normality atom as a condition. In the second part of the convention, if a normality atom appears as condition in a rule, it is unique to that rule. No other rule in the knowledgebase has the same normality atom as a condition. However, multiple rules in the knowledgebase can have the same negated normality atom as head. Then, in the third part of the convention, we also assume that if there is a rule with a negated normality atom as head, then there is another rule with that atom in the tail. This ensures that for a rule to be an exception, there is an ordinary rule to which it is an exception.

Example 16. For ordinary atoms $\mathcal{Q} = \{\mathtt{b}, \mathtt{p}, \mathtt{f}\}$ and normality atoms $\mathcal{N} = \{\mathtt{n}_1, \mathtt{n}_2\}$, a rule with normality atom as a condition is $\mathtt{b} \wedge \mathtt{n}_1 \rightarrow \mathtt{f}$, a rule with normality atom as a head is $\mathtt{p} \rightarrow \neg \mathtt{n}_1$, and a rule with normality atoms as a condition and as a head is $\mathtt{p} \wedge \mathtt{n}_2 \rightarrow \neg \mathtt{n}_1$.

We quantify the probability of a normality atom in terms of the unique rule that contains it as an antecedent: For a defeasible rule $\beta_1 \wedge \ldots \wedge \beta_n \wedge \gamma \rightarrow \alpha$ with normality atom γ, we assume $\gamma \equiv \beta_1 \wedge \ldots \wedge \beta_n \wedge \alpha$, and so $P(\gamma)$ is $P(\beta_1 \wedge \ldots \wedge \beta_n \wedge \alpha)$.

Example 17. Consider probability distribution P, with the rules in Δ being $\mathtt{b} \wedge \mathtt{n}_1 \rightarrow \mathtt{f}$ and $\mathtt{p} \wedge \mathtt{n}_2 \rightarrow \neg \mathtt{f}$. So we assume that $\mathtt{n}_1 \equiv \mathtt{b} \wedge \mathtt{f}$ and $\mathtt{n}_2 \equiv \mathtt{p} \wedge \neg \mathtt{f}$, and therefore, we have $P(\mathtt{n}_1) = P(\mathtt{b} \wedge \mathtt{f})$ and $P(\mathtt{n}_2) = P(\mathtt{p} \wedge \neg \mathtt{f})$.

$\langle \mathtt{b}, \mathtt{p}, \mathtt{f}, \mathtt{n}_1, \mathtt{n}_2 \rangle$	11001	10110	10000
P	0.01	0.95	0.04

The modelling of normality atoms contributes to the solution for addressing Goal 1. It provides a way of quantifying normality in defeasible rules in a probabilistic context, thereby ensuring the probability-based measures of argument strength that we will be proposing can be used with arguments involving normality atoms.

6. Conditional probabilities in argumentation

In this section, we consider how we can analyze the relationship between a pair of formulae, where one is a premise and the other is a claim, using conditional probabilities. This will address Goal 2, and in turn will help us to address Goals 3 and 4 in the subsequent sections.

If we have a probability distribution P, then the following are all the conditional probability statements involving both formulae ϕ and ψ, where $\phi, \psi \in \mathcal{B}(\mathcal{A})$, and their negations (i.e. we are not interested in probability statements such as $P(\phi|\phi)$, $P(\phi|\neg\phi)$ or $P(\neg\psi|\psi)$ which only involve one of the formula and/or its negation):

$(a)\ P(\phi|\psi)$ $(b)\ P(\phi|\neg\psi)$ $(c)\ P(\psi|\phi)$ $(d)\ P(\psi|\neg\phi)$
$(e)\ P(\neg\phi|\psi)$ $(f)\ P(\neg\phi|\neg\psi)$ $(g)\ P(\neg\psi|\phi)$ $(h)\ P(\neg\psi|\neg\phi)$

We consider these conditional probability statements as follows where we regard ψ as a premise and ϕ as claim. So we are concerned with how much ψ tells us about ϕ. Though, we do not assume that $\{\psi\} \vdash \phi$ holds. We will refer to (a) to (d) as our **key conditional probabilities** which we will use to analyze notions of sufficiency and necessity of the premise for the claim as discussed in the motivating examples in Section 1.1 and we discuss further below. We can obtain (e) to (h) from the key conditional probabilities as follows: $P(\neg\phi|\psi) = 1 - P(\phi|\psi)$, $P(\neg\phi|\neg\psi) = 1 - P(\phi|\neg\psi)$, $P(\neg\psi|\phi) = 1 - P(\psi|\phi)$, and $P(\neg\psi|\neg\phi) = 1 - P(\psi|\neg\phi)$. Hence, if we have the key conditional probabilities we can immediately obtain (e) to (h).

Sufficiency of premise for claim The conditional probability $P(\phi|\psi)$ measures the sufficiency of premise ψ for claim ϕ. As $P(\phi|\psi)$ rises, the probability that ϕ holds when ψ holds rises. If $P(\phi|\psi) = 0.5$, then the probability of ϕ holding given ψ holds is equal to the probability of $\neg\phi$ holding given ψ holds.

Sufficiency of complementary premise for claim The conditional probability $P(\phi|\neg\psi)$ measures the sufficiency of the complementary premise for the claim. As $\neg\psi$ constitutes alternative premises (i.e. alternative reasons) to ψ, then $P(\phi|\neg\psi)$ captures alternative reasons for the claim. Note, it is possible for $P(\phi|\psi) > 0.5$ and $P(\phi|\neg\psi) > 0.5$ to both hold.

Necessity of premise for claim The conditional probability $P(\psi|\phi)$ measures the necessity of premise ψ for claim ϕ. In other words, how important ϕ is for ψ. This captures the degree to which premise is an explanation for the claim. So if we require an explanation for the claim, rather than a prediction from the premise, we are more interested in a higher value for $P(\psi|\phi)$. Note, this is a specific view on the notion explanation that only considers the premises and claim, and not the broader context of explanation (see [16] for a review of explanation in argumentation).

Necessity of premise for complementary claim The conditional probability $P(\psi|\neg\phi)$ measures the necessity of premise ψ for complementary claim $\neg\phi$. This captures the degree to which the premise is an explanation for the complementary claim. So if we require an explanation for the claim as opposed to the complementary claim, we want a lower value for $P(\psi|\neg\phi)$.

The above discussion of sufficiency and necessity shows that there is more to understanding the relationship between premises and a claim than represented by the conditional probability $P(\phi|\psi)$. So we regard the conditional probabilities (i.e. (a) to (d)) as providing valuable information about the relationship between a premise ψ and a claim ϕ where $\phi, \psi \in \mathcal{B}(\mathcal{A})$.

Example 18. For the signature $\langle \mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d} \rangle$, where $\mathtt{a}$ is a claim, consider the following probability distribution. For this probability distribution, $\mathtt{a}$ holds when exactly one of $\mathtt{b}$, $\mathtt{c}$ or $\mathtt{d}$ hold. Also, whenever any of $\mathtt{b}$, $\mathtt{c}$ or $\mathtt{d}$ hold, then $\mathtt{a}$ is more likely to hold than not.

	1100	1010	1001	0100	0010	0001
P	0.4	0.2	0.2	0.05	0.05	0.1

So the key conditional probabilities are: $P(\mathtt{a}|\mathtt{b}) = 0.88$, so the sufficiency of the premise for the claim is high; $P(\mathtt{b}|\mathtt{a}) = 0.5$, so the necessity of the premise for the claim is neither high nor low; $P(\mathtt{a}|\neg\mathtt{b}) = 0.36$, so the sufficiency of the complementary premise for the claim is quite low; and $P(\mathtt{b}|\neg\mathtt{a}) = 0.25$, so the necessity of the premises for the complementary claim is low.

We now consider some properties of the key conditional probability. When the probability distribution is uniform, then the key conditional probability statements coincide, and hence indicate neutral sufficiency and necessity by both the evidence and its negation for the claim or negated claim.

Proposition 3. *For a probability distribution P, premise ψ, and claim ϕ, if P is uniform, then $P(\phi|\psi) = P(\psi|\phi) = P(\phi|\neg\psi) = P(\psi|\neg\phi) = 0.5$.*

The following result shows that both ψ and $\neg\psi$ can be completely sufficient reasons for ϕ, and so for any ψ and ψ', both ψ and ψ' can be completely sufficient reasons for ϕ.

Proposition 4. *For a probability distribution P, premise ψ and claim ϕ, $P(\phi|\psi) = 1$ and $P(\phi|\neg\psi) = 1$ iff $P(\phi \wedge \psi) + P(\phi \wedge \neg\psi) = 1$ and $P(\phi \wedge \psi) > 0$ and $P(\phi \wedge \neg\psi) > 0$.*

The next result shows that to have a premise that is completely necessary and completely sufficient when the probability of the conjunction of the premise and claim is equal to either of them individually.

Proposition 5. *For a probability distribution P, premise ϕ and claim ψ, $P(\phi|\psi) = 1$ and $P(\psi|\phi) = 1$ iff $P(\phi \wedge \psi) = P(\phi)$ and $P(\phi \wedge \psi) = P(\psi)$.*

The previous two results taken together show that we cannot have both $P(\phi|\psi) = 1$ and $P(\psi|\phi) = 1$ holding when $P(\psi|\neg\phi) > 0$ holds. In other words, we cannot have a reason being completely sufficient and necessary, and at the same time, have another reason with non-zero sufficiency for the claim.

Further observations come from logical relationships between the premises and the claim of an argument as captured next.

Proposition 6. *Let P be a probability distribution, and $\phi, \psi, \psi_1, \psi_2 \in \mathcal{B}(\mathcal{A})$. (1) $P(\phi|\psi) = 1$ iff $\{\psi\} \vdash \phi$; (2) $P(\phi|\psi) = 0$ iff $\{\psi\} \vdash \neg\phi$; And (3) $P(\phi_1|\psi) \leq P(\phi_2|\psi)$ if $\{\phi_1\} \vdash \phi_2$.*

Note, the converse of the third point in the above proposition does not necessarily hold (i.e. $\{\psi_1\} \vdash \psi_2$ does not imply $P(\phi|\psi_1) \leq P(\phi|\psi_2)$) as illustrated by the following example.

Example 19. For the signature $\langle \mathtt{a}, \mathtt{b}, \mathtt{c} \rangle$, let $P(110) = 0.5$ and $P(001) = 0.5$. So $P(\mathtt{a}|\mathtt{b}) = 1$ whereas $P(\mathtt{a}|\mathtt{b} \vee \mathtt{c}) = 0.5$.

Whilst $P(\neg\psi|\neg\phi)$ is the contrapositive of $P(\phi|\psi)$, it is not necessarily the case that $P(\neg\psi|\neg\phi) = P(\phi|\psi)$ as illustrated by the following counterexample.

Example 20. For the signature $\langle \mathtt{a}, \mathtt{b} \rangle$, $P(11) = 0.4$, $P(10) = 0.3$, $P(01) = 0.2$, and $P(00) = 0.1$, hence $P(\mathtt{a}|\mathtt{b}) = 0.66$ and $P(\neg\mathtt{b}|\neg\mathtt{a}) = 0.25$, and so $P(\neg\mathtt{b}|\neg\mathtt{a}) \neq P(\mathtt{a}|\mathtt{b})$.

Whilst we have considered premises and claims in general in this section, we can focus on argumentation by treating the conjunction of the support of the argument as the premise and the claim of the argument as the claim. So for probabilistic argumentation, we can analyze an argument in terms of the claim conditioned on the premise (as considered in for example [50,62,46,51,34,18]). But as indicated by the discussion in this section, if we just do this, then we may get a misleading picture.

Furthermore, we may need to consider the relationship between the probability of a claim and the probability of a claim given a premise. For instance, for a premise $\mathtt{b}$ and a claim $\mathtt{a}$, we may have a high value for $P(\mathtt{a}|\mathtt{b})$, indicating that $\mathtt{b}$ is good evidence for $\mathtt{a}$. Yet we may have an even higher value for $P(\mathtt{a})$, thereby indicating that conditioning on $\mathtt{b}$ has reduced the probability of the claim (e.g. for a signature $\langle \mathtt{a}, \mathtt{b} \rangle$, $P(11) = 0.5$, $P(10) = 0.3$, $P(01) = 0.2$, and $P(00) = 0$ is an example since $P(\mathtt{a}|\mathtt{b}) = 0.71$ and $P(\mathtt{a}) = 0.8$). In this case, we might not want to consider the premises as good evidence for the claim. Similarly, if $P(\mathtt{a}|\mathtt{b}) = P(\mathtt{a})$, then $P(\mathtt{a})$ and $P(\mathtt{b})$ are independent, and again we might not want to consider the premises as good evidence for the claim.

So there are multiple issues including different kinds of conditional probabilities that we can consider. In particular, if we focus on the key conditional probabilities, then we obtain insights into the necessity and sufficiency of the evidence for the claim. This section therefore addresses Goal 2. However, there are some options for how we can use them to measure the strength of arguments. We consider this question in the following sections.

7. Probability-based measures of argument strength

Argument strength is a commonly used term. But, since the term is used in multiple senses, we need to investigate definitions for it. A probabilistic perspective can give us some options for this. So in this section, we will address Goal 3. As we see from the investigation of conditional probabilities in the previous section, there are four key conditional probabilities that we may wish to take into account in definitions for probability-based measures of argument strength.

In the rest of this paper, we will focus on arguments that involve relationships between observations. We will assume that all the atoms, apart from the normality atoms, are **observations**. These are atoms that can ultimately be verified as true or false, though at any specific time, there may be uncertainty about which observations are true or false. Examples of observations include $\mathtt{b}$ = *bird*, $\mathtt{d}$ = *duck*, $\mathtt{p}$ = *penguin*, $\mathtt{e}$ = *eagle*, and $\mathtt{f}$ = *capable-of-flying*. Recall that in Section 5, we partitioned the set of atoms $\mathcal{A}$ into ordinary atoms $\mathcal{Q}$ and normality atoms $\mathcal{N}$, and so we will assume that the ordinary atoms are observations.

Once we decide on the atoms for representing observations, we can obtain a probability distribution by assuming a **frame of reference**. For instance, for birds flying, we can consider a specific zoo, or an aviary, or a specific location (e.g. our back garden, or a specific park), or type of location (e.g. all back gardens in London), etc. This will help identify appropriate probability distributions whether this done from a subjective or frequentist perspective.

Given an argument A, the **evidence** in A, denoted $\mathsf{E}(A)$, is the conjunction of the observations in the support of A that are not normality atoms. (i.e. $\mathsf{E}(A) = \phi_1 \wedge \ldots \wedge \phi_k$ when $\mathsf{Facts}(A) \cap \mathcal{L}(\mathcal{Q}) = \{\phi_1, \ldots, \phi_k\} \neq \emptyset$ and $\mathsf{E}(A) = \top$ when $\mathsf{Facts}(A) \cap \mathcal{L}(\mathcal{Q}) = \emptyset$). So we will consider strength in terms how the evidence in the premises supports the claim.

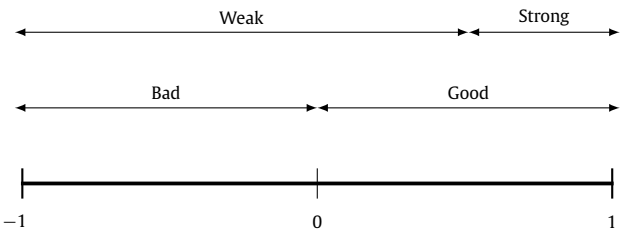

Fig. 2. Terms for describing arguments according to a measure of strength. The threshold that makes the boundary between strong and weak is a value in the unit interval.

For an argument A, we let $S(A)$ be an assignment in the $[-1, 1]$ interval. If $S(A) > 0$, then the argument is good to some degree and in some respect, and if $S(A) < 0$, then the argument is bad to some degree and in some respect. As $S(A)$ rises above (respectively below) zero, then the support of the argument gives an increasingly good (respectively bad) argument.

We may wish to clarify what it means to say that an argument is strong. We can set a threshold $\tau \in (0, 1]$ so that if $S(A) \geq \tau$, then A is a strong argument, otherwise it is a weak argument. So an argument has to be good for it to be strong, and all bad arguments are weak arguments. Some good arguments may also be weak arguments depending on the choice of threshold. We summarize these concepts in Fig. 2.

Since we are concerned with probability-based measures of argument strength, we will consider strength measures S_P where P is a probability distribution. Furthermore, we can be more specific in what we mean by the measures: If $S_P(A) > 0$, then the evidence in the support provides a good reason for the claim to some degree and in some respect according to P, if $S(A) = 0$, then the support is neither a good nor bad reason in some respect according to P, and if $S(A) < 0$, then the evidence in the support provides a bad reason for the claim to some degree and in some respect according to P. This still leaves open in what respects a probability-based measure actually measures argument strength. In order to investigate this, we will consider a set of properties for characterizing probability-based measures of argument strength. We will discuss the appropriateness of the properties in each set in the text below the definition for the set.

Equivalence properties These capture when arguments have the same strength. For EQ1, it is when the evidence in the arguments are logically equivalent and the claims in the arguments are logically equivalent; and for EQ2, it is when one argument is the contrapositive of the other.

(EQ1) If $\mathsf{E}(A) \equiv \mathsf{E}(B)$ and $\mathsf{C}(A) \equiv \mathsf{C}(B)$, then $S(A) = S(B)$
(EQ2) If $\mathsf{E}(A) \equiv \neg\mathsf{C}(B)$ and $\mathsf{C}(A) \equiv \neg\mathsf{E}(B)$, then $S(A) = S(B)$

In most applications, EQ1 is a natural property to assume. A possible exception is if we want to take the syntactic complexity of the formulae into account, as it might have more impact with an audience, in which case we may regard an argument with syntactically simpler premises to be a stronger argument. EQ2 is more controversial. In a classical logic setting, this would be a natural property to assume. But when considering uncertainty, the connection between evidence and a claim is not necessarily assumed to be equal to that of its contrapositive.

Maximum strength properties These capture when arguments have maximum strength. For MAX1, this is when the claim is logically equivalent to the evidence, and for MAX2, this is when the belief in the conjunction of the evidence and claim is 1.

(MAX1) If $\mathsf{E}(A) \equiv \mathsf{C}(A)$, then $S(A) = 1$
(MAX2) If $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 1$, then $S(A) = 1$

In some applications, MAX1 would be a natural property to assume. However, for some applications, equivalence of evidence and claim might be regarded as insufficient to ensure the maximum strength when for instance strength is to reflect how well the evidence explains the claim. In many applications, MAX2 would be a natural property to assume as it reflects the situation where evidence and claim are both certain, and so in this sense, the argument has maximum strength.

Zero strength properties These capture when arguments have zero strength: For ZERO1, this is when there is a uniform distribution, so there is no material relationship between the evidence and claim; For ZERO2, this is when the sufficiency of the evidence for the claim is the same as the probability of the claim; For ZERO3, this is when the necessity of the evidence for the claim is equal to the necessity of the evidence for the negated claim; For ZERO4, this is when the sufficiency of the evidence for the claim is equal to sufficiency of the evidence for the complementary claim; And for ZERO5, this is when the sufficiency of the evidence for the claim is equal to the sufficiency of the negation of the evidence for the claim.

(ZERO1) If P is uniform, then $S(A) = 0$
(ZERO2) If $P(\mathsf{C}(A)|\mathsf{E}(A)) = P(\mathsf{C}(A))$, then $S(A) = 0$
(ZERO3) If $P(\mathsf{E}(A)|\mathsf{C}(A)) = P(\mathsf{E}(A)|\neg\mathsf{C}(A))$, then $S(A) = 0$
(ZERO4) If $P(\mathsf{C}(A)|\mathsf{E}(A)) = P(\neg\mathsf{C}(A)|\mathsf{E}(A))$, then $S(A) = 0$
(ZERO5) If $P(\mathsf{E}(A)|\mathsf{C}(A)) = P(\neg\mathsf{E}(A)|\mathsf{C}(A))$, then $S(A) = 0$

In most applications, ZERO1 is a natural property to assume since a uniform distribution gives no information, and so its effect is neither positive nor negative on the strength. ZERO2 is an important property to consider for the strength being neither positive nor negative when the evidence does not change the belief in the claim. Whether to adopt ZERO3, ZERO4, or ZERO5, depends on what we want to measure and each gives a useful notion of zero strength and so help us to characterize different measures. For instance, ZERO4 defines zero strength as when the probability of the claim given the evidence is 0.5, and ZERO5 defines zero strength as when the probability of the evidence given the claim is 0.5.

Positive strength properties These capture when arguments have positive strength: For POS1, the strength is positive only if the sufficiency of the claim given the evidence is greater than zero; For POS2, the strength is positive only if the sufficiency of the claim given the evidence is greater than 0.5; For POS3, the strength is positive only if the necessity of the claim given the evidence is greater than 0.5; For POS4, if the sufficiency of the evidence for the claim is greater than the probability of the claim, then the strength is positive; For POS5, the sufficiency of the claim given the evidence is greater than zero, the strength is positive; And for POS6, the necessity of the claim given the evidence is greater than zero, the strength is positive.

(POS1) If $S(A) > 0$, then $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0$
(POS2) If $S(A) > 0$, then $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0.5$
(POS3) If $S(A) > 0$, then $P(\mathsf{E}(A)|\mathsf{C}(A)) > 0.5$
(POS4) If $P(\mathsf{C}(A)|\mathsf{E}(A)) > P(\mathsf{C}(A)$, then $S(A) > 0$
(POS5) If $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0$, then $S(A) > 0$
(POS6) If $P(\mathsf{E}(A)|\mathsf{C}(A)) > 0$, then $S(A) > 0$

For most applications, POS1 would be a natural property. It is difficult to conceive of a strength measure for which this would fail. Whilst a threshold of 0.5 might seem reasonable for POS2 and POS3, they are actually quite difficult to satisfy, and might not be appropriate for many applications. POS4 is an important property to consider as it captures an important aspect of argument strength (viz. that of prediction) and it is the positive counterpart to ZERO2. The conditions of POS5 and POS6 are easily satisfied, and hence, it is arguable whether they should imply that the argument strength is positive.

Negative strength properties These capture when arguments that have negative strength. For NEG1, if the joint probability of the claim and evidence is zero, then the strength is not positive; For NEG2, if the sufficiency of the claim given the evidence is zero, then the strength is not positive; For NEG3, if the strength is negative, then the sufficiency of the claim given the evidence is zero; For NEG4, if the strength is negative, then the necessity of the claim given the evidence is zero; And for NEG5, if the sufficiency of the evidence for the claim is less than the probability of the claim, then the strength is negative.

(NEG1) If $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$, then $S(A) \leq 0$
(NEG2) If $P(\mathsf{C}(A)|\mathsf{E}(A)) = 0$, then $S(A) \leq 0$
(NEG3) If $S(A) < 0$, then $P(\mathsf{C}(A)|\mathsf{E}(A)) = 0$
(NEG4) If $S(A) < 0$, then $P(\mathsf{E}(A)|\mathsf{C}(A)) = 0$
(NEG5) If $P(\mathsf{C}(A)|\mathsf{E}(A)) < P(\mathsf{C}(A)$, then $S(A) < 0$

For most applications, NEG1 and NEG2 would be natural properties to assume. In contrast, NEG3 and NEG4 would often be too strict, since we would not necessary want negative strength to imply that the claim conditioned on the evidence or vice versa would be zero. NEG5 is also an important property to consider as it captures an important aspect of argument strength (viz. that of prediction) and it is the negative counterpart to ZERO2 and POS4.

Counterargument properties These capture when one argument attacks another argument. For INC1, this is when the arguments rebut, and for INC2, this is when one of the arguments undercuts the other, and as a result in both cases the strength of one of arguments is below zero.

(INC1) If $\{\mathsf{C}(A), \mathsf{C}(B)\} \vdash \bot$, then $S(A) < 0$ or $S(B) < 0$
(INC2) If $\{\mathsf{C}(A), \mathsf{E}(B)\} \vdash \bot$, then $S(A) < 0$ or $S(B) < 0$

Both INC1 and INC2 seem like natural properties at first sight. If two arguments rebutting claims, or the claim of one undercuts the other, then it would seem reasonable that one argument is not correct and therefore the strength is below zero. But, this conflates strength and acceptability, and so we may choose to reject these properties if we want to say that an argument can have positive strength in some sense even if we have a logical or dialectical criterion for rejecting it.

We will use the characterization properties presented in this section (which we have presented to address Goal 3) to compare and contrast specific definitions for probability-based measures of argument strength in Section 9.

8. Some probability-based measures of argument strength

We now consider Goal 4 by proposing four specific probability-based measures of argument strength: S^1_P is the measure of sufficiency (of the evidence for the claim) relative to the negated evidence; S^2_P is the measure of necessity (of the evidence for the claim) relative to the negated claim; S^3_P is the measure of sufficiency (of the evidence for the claim) relative to the negated claim; And S^4_P is the measure of necessity (of the evidence for the claim) relative to the negated evidence. These measures are defined in terms of the key conditional probabilities that we discussed in Section 6.

Definition 11. For an argument A, and probability distribution P, the S^1_P, S^2_P, S^3_P, and S^4_P, **probability-based measures of argument strength** are defined as follows.

$$S^1_P(A) = \begin{cases} P(\mathsf{C}(A)|\mathsf{E}(A)) \text{ if } P(\mathsf{C}(A)|\neg\mathsf{E}(A)) \text{ is undefined} \\ -P(\mathsf{C}(A)|\neg\mathsf{E}(A)) \text{ if } P(\mathsf{C}(A)|\mathsf{E}(A)) \text{ is undefined} \\ P(\mathsf{C}(A)|\mathsf{E}(A)) - P(\mathsf{C}(A)|\neg\mathsf{E}(A)) \text{ otherwise} \end{cases} \quad (\textbf{relative sufficiency})$$

$$S^2_P(A) = \begin{cases} P(\mathsf{E}(A)|\mathsf{C}(A)) \text{ if } P(\mathsf{E}(A)|\neg\mathsf{C}(A)) \text{ is undefined} \\ -P(\mathsf{E}(A)|\neg\mathsf{C}(A)) \text{ if } P(\mathsf{E}(A)|\mathsf{C}(A)) \text{ is undefined} \\ P(\mathsf{E}(A)|\mathsf{C}(A)) - P(\mathsf{E}(A)|\neg\mathsf{C}(A)) \text{ otherwise} \end{cases} \quad (\textbf{relative necessity})$$

$$S^3_P(A) = \begin{cases} 0 \text{ if } P(\mathsf{C}(A)|\mathsf{E}(A)) \text{ and } P(\neg\mathsf{C}(A)|\mathsf{E}(A)) \text{ are undefined} \\ P(\mathsf{C}(A)|\mathsf{E}(A)) - P(\neg\mathsf{C}(A)|\mathsf{E}(A)) \text{ otherwise} \end{cases} \quad (\textbf{normalized sufficiency})$$

$$S^4_P(A) = \begin{cases} 0 \text{ if } P(\mathsf{E}(A)|\mathsf{C}(A)) \text{ and } P(\neg\mathsf{E}(A)|\mathsf{C}(A)) \text{ are undefined} \\ P(\mathsf{E}(A)|\mathsf{C}(A)) - P(\neg\mathsf{E}(A)|\mathsf{C}(A)) \text{ otherwise} \end{cases} \quad (\textbf{normalized necessity})$$

The first two of these measures have been adopted from specific confirmation measures: S^1_P is the measure of [14]; and S^2_P is the measure of [44]. The third measure maps the value for the probability of the claim conditioned on the evidence into the $[-1, 1]$ interval, and the fourth measure maps the value for the probability of the evidence conditioned on the claim into the $[-1, 1]$ interval.

In the above definition for each of S^1_P, S^2_P, S^3_P, and S^4_P, we need to consider the cases where one or other conditional probability is undefined. Note, for S^1_P, it is not possible for $P(\mathsf{C}(A)|\mathsf{E}(A))$ and $P(\mathsf{C}(A)|\neg\mathsf{E}(A))$ to be both undefined, and for S^2_P, it is not possible for $P(\mathsf{E}(A)|\mathsf{C}(A))$ and $P(\mathsf{E}(A)|\neg\mathsf{C}(A))$ to be both undefined. Whereas for S^3_P, $P(\mathsf{C}(A)|\mathsf{E}(A))$ and $P(\neg\mathsf{C}(A)|\mathsf{E}(A))$ are either both defined or both undefined, and for S^4_P, $P(\mathsf{E}(A)|\mathsf{C}(A))$ and $P(\neg\mathsf{E}(A)|\mathsf{C}(A))$ are either both defined or both undefined.

We might consider the application of these measures to an argument as follows: (S^1_P) For the relative sufficiency measure, we should only use arguments A where $S^1_P(A) > 0$ (if it is below zero, then the negation of the evidence is more likely to imply the claim). (S^2_P) For the relative necessity measure, when the number of instances of $\mathsf{E}(A)$ and $\neg\mathsf{E}(A)$ are similar, we then only want to use arguments A where $S^2_P(A) > 0$ (if it is below zero, then the evidence is more likely to be necessary for the negation of the claim). (S^3_P) For the normalized sufficiency measure, we should only use arguments A where $S^3_P(A) > 0$ (if it is below zero, then the evidence is more likely to imply the negated claim). (S^4_P) For the normalized necessity measure, when the number of instances of $\mathsf{E}(A)$ and $\neg\mathsf{E}(A)$ are similar, we then only want to use arguments A where $S^4_P(A) > 0$ (if it is below zero, then the negation of the evidence is more likely to be necessary for the claim).

Example 21. Consider the probability distribution P in the left table, with signature $\langle \mathtt{b}, \mathtt{p}, \mathtt{f} \rangle$, where $\mathtt{b}$ is *bird*, $\mathtt{p}$ is *penguin*, $\mathtt{f}$ is *capable of flying*, for a zoo with a large aviary.

$\langle \mathtt{b}, \mathtt{p}, \mathtt{f}, \mathtt{n} \rangle$	1100	1011	1000	0000
P	0.01	0.75	0.04	0.2

	$P(.)$	S^1_P	S^2_P	S^3_P	S^4_P
A_1	0.75	0.94	0.80	0.88	1.00
A_2	0.01	0.76	0.04	1.00	−0.92
A_3	0.01	0.76	0.04	1.00	−0.92

For the arguments $A_1 = \langle \{\mathtt{b}, \mathtt{b} \wedge \mathtt{n} \rightarrow \mathtt{f}\}, \mathtt{f} \rangle$, $A_2 = \langle \{\mathtt{p}, \mathtt{p} \rightarrow \neg\mathtt{f}\}, \neg\mathtt{f} \rangle$, and $A_3 = \langle \{\mathtt{p}, \mathtt{p} \rightarrow \neg\mathtt{n}\}, \neg\mathtt{n} \rangle$. The strengths are given in the right table. So A_1 has a high probability of acceptability, and good scores for all strength scores, and A_2 has a low probability of acceptability, good scores for the sufficiency scores, and low scores for necessity scores (but as penguins are a small subset of the flying things, and so this may be regarded as acceptable). Note, A_3 has the same values as A_2 because $\mathtt{n}$

$\equiv \mathtt{b} \wedge \mathtt{f}$, and $\neg\mathtt{n} \equiv \neg\mathtt{b} \vee \neg\mathtt{f}$. And since, the probability distribution reflects that all penguins are birds, $P(\neg\mathtt{n})$ is the same as $P(\neg\mathtt{f})$.

In the above example, the importance of the low scores for necessity does depend on what the argument is being used for. Is it explaining the penguin's ability to fly, or is it explaining what does not fly? In the latter case, low necessity indicates that a good explanation has not been found even though the evidence is sufficient for inferring the claim.

Example 22. We now consider an example adapted from [18] which was in turn adapted from [24]. Let $\mathcal{S} = \langle \mathtt{c}, \mathtt{i}, \mathtt{s}, \mathtt{f} \rangle$. The situation concerns a chicken (denoted $\mathtt{c}$) in a coup, incubating an egg (denoted $\mathtt{i}$), being scared (denoted $\mathtt{s}$), and flying (denoted $\mathtt{f}$). The following probability distribution is based on counting the number of nights where the chicken is incubating, the chicken appears scared (perhaps because it has heard a noise that it thinks may be a fox) during the night, and the chicken flying at least once during the night.

	1111	1110	1101	1100	1011	1010	1001	1000
P	0.001	0.004	0.000	0.045	0.094	0.001	0.005	0.850

Next we consider the following arguments and their strength measures using the above probability distribution.

	S^1_P	S^2_P	S^3_P	S^4_P
$A_1 = \langle \{\mathtt{c}, \mathtt{c} \rightarrow \neg\mathtt{f}\}, \neg\mathtt{f} \rangle$	0.900	0.000	0.800	1.000
$A_2 = \langle \{\mathtt{c}, \mathtt{s}, \mathtt{c} \wedge \mathtt{s} \rightarrow \mathtt{f}\}, \mathtt{f} \rangle$	0.944	0.944	0.900	0.900
$A_3 = \langle \{\mathtt{c}, \mathtt{s}, \mathtt{i}, \mathtt{c} \wedge \mathtt{s} \wedge \mathtt{i} \rightarrow \mathtt{f}\}, \mathtt{f} \rangle$	0.101	0.006	−0.600	−0.980

We explain these evaluations as follows: A_1 is in general a strong argument as it has a high value for S^1_P, as in most cases, the chicken is not flying, but the value for S^2_P is near zero, and so knowing that it is a chicken is not necessary for determining whether it is flying; A_2 is a strong argument as it has high values for sufficiency and necessity of the evidence for the claim and so knowing that the chicken is scared is good evidence for the claim; But A_3 is a bad argument as the extra evidence that the chicken is nesting substantially weakens the strength of the argument.

To illustrate how each of the four measures identifies something different and useful about the relationship of the evidence to the claim, we consider the following example.

Example 23. Imagine a clinic where data is collected on the proportion of patients who have specific symptom $\mathtt{s}$ and the proportion of patients who have specific disease $\mathtt{d}$. So we can consider the argument $A = \langle \{\mathtt{s}, \mathtt{s} \rightarrow \mathtt{d}\}, \mathtt{d} \rangle$. Let $S = \langle \mathtt{d}, \mathtt{s} \rangle$ and below are some possible probability distributions.

	11	10	01	00	S^1_P	S^2_P	S^3_P	S^4_P
P_0	0.700	0.010	0.010	0.280	0.952	0.952	0.972	0.972
P_1	0.690	0.301	0.001	0.008	0.024	0.585	0.997	0.393
P_2	0.790	0.001	0.201	0.008	0.686	0.037	0.594	0.997
P_3	0.126	0.067	0.196	0.611	0.292	0.410	−0.217	0.306
P_4	0.126	0.196	0.067	0.611	0.410	0.292	0.306	−0.217
P_5	0.584	0.259	0.151	0.006	−0.183	−0.269	0.589	0.386
P_6	0.030	0.335	0.438	0.197	−0.566	−0.608	−0.872	−0.836

We can view each of these probability distributions as reflecting a scenario. We explain these scenarios and the resulting strength measures as follows.

- P_0 Over 2/3 of patients have the symptom and the disease, and nearly 1/3 of patients have neither the symptom nor the disease. All scores are high.
- P_1 Most patients have the disease and about two-thirds have the symptom. All score are quite positive except S^1_P which is low because the sufficiency of $\mathtt{s}$ for $\mathtt{d}$ is similar to the sufficiency of $\neg\mathtt{s}$ for $\mathtt{d}$.
- P_2 Most patients have the symptom and the disease (i.e. most mass is on $\mathtt{d} \wedge \mathtt{s}$). All scores are quite high except S^2_P because the necessity of $\mathtt{s}$ for $\mathtt{d}$ and for $\neg\mathtt{d}$ are very similar.
- P_3 Just over half of patients have neither the symptom nor the disease. All scores are positive except S^3_P which is negative because the sufficiency of $\mathtt{s}$ for $\neg\mathtt{d}$ is greater than the sufficiency of $\mathtt{s}$ for $\mathtt{d}$.
- P_4 Just over half of patients have neither the symptom nor the disease. All scores are positive except S^4_P which is negative because the necessity of $\neg\mathtt{s}$ for $\mathtt{d}$ is greater than the necessity of $\mathtt{s}$ for $\mathtt{d}$.
- P_5 Over three-quarters of patients have the disease, and over half have the symptom and disease. Both the S^1_P and S^2_P scores are negative because the sufficiency of $\mathtt{s}$ for $\mathtt{d}$ is lower than that of $\neg\mathtt{s}$ for $\mathtt{d}$ and the necessity of $\mathtt{s}$ for $\mathtt{d}$ is lower than that of $\mathtt{s}$ for $\neg\mathtt{d}$.

- P_6 Most patients who have the disease do not have the symptom, and most patients who have the symptom do not have the disease. Hence, all scores are negative.

So apart for P_0, for each of the above probability distributions suggest that the argument A is weak in one or more respects.

In this section, we have presented four probability-based measures for argument strength as a proposal to address Goal 4. How we should consider a trade-off of scores depends on the application. For instance, if we have an argument, and we want to know to what degree that the claim holds, we might be more interested in the sufficiency measures whereas if we want to know to what the premises are a good explanation of the claim, we might be more interested in the necessity measures.

9. Properties of probability-based measures of strength

We now investigate properties of the four probability-based measures of strength that we introduced in the previous section and thereby continue to address Goal 4. We start by considering the conditions under which we get the minimum and maximum strength for each measure as illustrated in the following example.

Example 24. The following table lists the extreme probability distributions for an argument A, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$, and the corresponding probability-based measures of strength.

	$\phi \wedge \psi$	$\phi \wedge \neg\psi$	$\neg\phi \wedge \psi$	$\neg\phi \wedge \neg\psi$	S^1_P	S^2_P	S^3_P	S^4_P
P_1	1	0	0	0	1	1	1	1
P_2	0	1	0	0	-1	0	0	-1
P_3	0	0	1	0	0	-1	-1	0
P_4	0	0	0	1	0	0	0	0

We generalize the above example in the form of the following results.

Proposition 7. *For any argument A, and probability distribution P,*

1. $S^1_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$
2. $S^1_P(A) = -1$ *iff* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$
3. $S^2_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$
4. $S^2_P(A) = -1$ *iff* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$
5. $S^3_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$
6. $S^3_P(A) = -1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$
7. $S^4_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$
8. $S^4_P(A) = -1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) \neq 0$

Corollary 1. *For any argument A, and probability distribution P,*

1. $S^1_P(A) = 1$ *iff* $S^2_P(A) = 1$
2. $S^3_P(A) = 1$ *if* $S^1_P(A) = 1$ *or* $S^2_P(A) = 1$
3. $S^3_P(A) = -1$ *if* $S^2_P(A) = -1$
4. $S^4_P(A) = 1$ *if* $S^1_P(A) = 1$ *or* $S^2_P(A) = 1$
5. $S^4_P(A) = -1$ *if* $S^1_P(A) = -1$

The next result captures which measures agree in the direction of strength (i.e. which measure are positive or negative for the same argument and probability distribution). We introduce the notion of independence which means that for measures S^i_P and S^j_P, a positive score for $S^i_P(A)$ does not imply a positive score for $S^j_P(A)$, and a negative score for $S^i_P(A)$ does not imply a negative score for $S^j_P(A)$. So we say that S^i_P is **independent** of S^j_P when there is an argument A and probability distribution P such that $S^i_P(A) > 0$ (respectively $S^i_P(A) < 0$) and $S^j_P(A) < 0$ (respectively $S^j_P(A) > 0$). Furthermore, S^i_P is not independent of S^j_P when for any argument A, and probability distribution P, $S^i_P(A)\#0$ iff $S^j_P(A)\#0$ for $\# \in \{<, \leq, =, \geq, >\}$.

We do a pairwise comparison according to independence of the four measures in the following result. Showing that two measure are independent of each other is one way of demonstrating that they measure different aspects of strength of an argument.

Table 1
The table captures satisfaction ✓, or non-satisfaction ×, of the characterization property. The results are given in the Appendix in Propositions 10 to 31.

	S^1_P	S^2_P	S^3_P	S^4_P
EQ1	✓	✓	✓	✓
EQ2	×	×	×	×
MAX1	✓	×	✓	×
MAX2	✓	✓	✓	✓
ZERO1	✓	✓	✓	✓
ZERO2	✓	✓	×	×
ZERO3	✓	✓	×	×
ZERO4	×	×	✓	×
ZERO5	×	×	×	✓
POS1	✓	✓	✓	✓
POS2	×	×	✓	×
POS3	×	×	×	✓
POS4	✓	×	×	×
POS5	×	×	×	×
POS6	×	×	×	×
NEG1	✓	✓	✓	✓
NEG2	✓	✓	✓	✓
NEG3	×	×	×	×
NEG4	×	×	×	×
NEG5	✓	✓	×	×
INC1	×	×	×	×
INC2	×	×	×	×

Proposition 8. *For measures S^1_P, S^2_P, S^3_P, and S^4_P, S^1_P is not independent of S^2_P, S^1_P is independent of S^3_P, S^1_P is independent of S^4_P, S^2_P is independent of S^3_P, S^2_P is independent of S^4_P, and S^3_P is independent of S^4_P.*

Unless we allow some of the key conditional probabilities to be undefined, we are not able to obtain a probability distribution such that all measures assign 1 (i.e. for an argument A, there is no probability distribution P, such that $S^i_P(A) = 1$ for all $i \in \{1, 2, 3, 4\}$). However, we are able to get arbitrarily close to 1 (i.e. all the measures give the same value that is close to 1) as shown in the next result. For this, a probability distribution P is a **non-zero probability distribution** iff for all $m \in \mathcal{M}(\mathcal{A})$, $P(m) \neq 0$.

Proposition 9. *For any argument A, for any $\delta \in (0, 1]$, there is a non-zero probability distribution P such that $S^i_P(A) = 1 - \delta$ for all $i \in \{1, 2, 3, 4\}$.*

We now consider the general properties given in Section 9 for the four measures. We give each result in the appendix (Propositions 10 to 31) and summarize these results in Table 1.

As can be seen in Table 1, none of the measures satisfy all the properties. Five of the properties are satisfied by all measures (viz. EQ1, MAX2, ZERO1, POS1, NEG1, and NEG2). These are the properties that we described as being natural to assume for most applications when we introduced them in Section 7. A further nine properties are satisfied by one or two of the four measures (viz. MAX1, ZERO2, ZERO3, ZERO4, ZERO5, POS2, POS3, POS4, and NEG5). Each of these captures an important dimension of argument strength, and so tells us something about an aspect of argument strength, but it is not the case that all measures need to satisfy them. Rather we can view these properties as delineating important features of that we want for some of our measures, or equivalently, we can view them as characterizing aspects of each measure. Finally, seven of the properties are not satisfied by any of the four measures (viz. EQ2, POS5, POS6, NEG3, NEG4, INC1 and INC2). Some of these properties seem to capture important aspects of argument strength, in particular EQ2, INC1, and INC2, and it would be worthwhile to seek measures that do satisfy one or more of these three properties, whereas for the remaining properties, we need further investigation to determine whether indeed they are not desirable for any measure of argument strength given the discussion of them in Section 7.

To conclude this section, we have shown how each of our four probability-based measures of argument strength provide a different perspective on an argument, and thereby contributed to addressing Goal 4. We have also shown how they are inter-related, and which general properties they satisfy. Note, we have only adopted two confirmation measures as measures of argument strength. There are a number of other confirmation measures that we could consider in future work (see Tentori *et al.* [59] for a review of confirmation measures).

10. Multiple defeasible rules

So for our examples of arguments, we have only used one defeasible rule in the support. We now consider the question of the strength of an argument with multiple defeasible rules in the premises. In effect, we will be using what we have gained so far from addressing Goals 1 to 4 to further enhance our solution to Goal 1. To illustrate some of our concerns, consider the following arguments where b denotes *bird*, w denotes *has wings*, y denotes *yellow*, f denotes *capable of flying*.

- $A_1 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$
- $A_2 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{w}, \mathtt{w} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$
- $A_3 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{y}, \mathtt{y} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$

Intuitively, A_1 is a reasonably strong argument since most birds have the capability to fly. But does A_2 have the same strength as A_1 since it starts from the same fact (i.e. bird) or is it stronger because it makes the intermediate point concerning having wings? And does A_3 have the same strength as A_1 or is it weaker because it makes the intermediate point (i.e. being yellow) that is irrelevant (and unlikely to be correct)? Assuming the probability distribution over the atoms b and f is the same for each argument, then the strength of each argument is the same since it is based on b and f. However, taking the defeasible rules into account, we might expect that A_2 is a strong argument since the premise *has wings* would be highly sufficient for *capable of flying*, whereas we might expect that A_3 is a weak argument since the premise *yellow* would not appear to be sufficient for *capable of flying*. To capture this, we consider how the assessment of the strength of an argument can depend on its intermediate steps.

Definition 12. Argument B is an **intermediate** of argument A iff $\mathsf{Rules}(B) \subseteq \mathsf{Rules}(A)$. Let $\mathsf{Intermediates}(A) = \{B \mid B \text{ is an intermediate of } A\}$.

Example 25. For $A_1 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{w}, \mathtt{w} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$, the intermediates are A_1, $B_1 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{w}\}, \mathtt{w}\rangle$, and $B_2 = \langle\{\mathtt{w}, \mathtt{w} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$. This is because $\mathsf{Rules}(B_1) \subseteq \mathsf{Rules}(A_1)$ and $\mathsf{Rules}(B_2) \subseteq \mathsf{Rules}(A_1)$. The atom w which is the support of B_2 but not A does not effect whether B_2 is an intermediate.

If B is a strict intermediate of A (i.e. $\mathsf{Rules}(B) \subset \mathsf{Rules}(A)$), and $\mathsf{C}(B) \neq \mathsf{C}(A)$, then there is defeasible rule $\psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi \in \mathsf{Support}(A)$ where $\mathsf{C}(B) \in \{\psi_1, \ldots, \psi_n\}$ (e.g. B_1 in Example 25). This is because arguments are minimal, and so if the claim of the intermediate differs from that of the argument, then it is also a condition in a defeasible rule. Also, if B is a strict intermediate of A, it is not necessarily the case that $\mathsf{Support}(B) \subset \mathsf{Support}(A)$ (e.g. B_2 in Example 25). This is because the definition of intermediates only requires the rules in the intermediate are in the original argument. In order to consider the intermediates in the derivation of a claim from its premises, we use the following definition that judges not just the argument but also the intermediates.

Definition 13. An argument A is **compositionally strong** with respect to strength measure S_P and threshold $\tau \in [-1, 1]$ iff for all $B \in \mathsf{Intermediates}(A)$, $S_P(B) \geq \tau$.

We now return to the arguments A_2 and A_3 from the introduction of this section, and analyze them in the following two examples.

Example 26. For $A_2 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{w}, \mathtt{w} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$, with the following probability distribution for a zoo, where $\mathcal{S} = \langle \mathtt{b}, \mathtt{w}, \mathtt{f}\rangle$, then A_2 and its strict intermediates have high strength.

$\langle \mathtt{b}, \mathtt{w}, \mathtt{f}\rangle$	111	110	001	000
P	0.09	0.01	0.02	0.88

	S^1_P	S^2_P	S^3_P	S^4_P
$\langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{w}, \mathtt{w} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$	0.88	0.81	0.80	0.64
$\langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{w}\}, \mathtt{w}\rangle$	1.00	1.00	1.00	1.00
$\langle\{\mathtt{w}, \mathtt{w} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$	0.88	0.81	0.80	0.64

Example 27. For $A_3 = \langle\{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{y}, \mathtt{y} \rightarrow \mathtt{f}\}, \mathtt{f}\rangle$, with the following probability distribution for a zoo, where $\mathcal{S} = \langle \mathtt{b}, \mathtt{y}, \mathtt{f}\rangle$, then A_3 has the same high strength as A_2 in the previous example (because the marginals involving b and f are the same) but low strength for the strict intermediates.

$\langle \mathtt{b}, \mathtt{y}, \mathtt{f} \rangle$	111	101	100	010	001	000
P	0.01	0.08	0.01	0.02	0.02	0.86

	S^1_P	S^2_P	S^3_P	S^4_P
$\langle \{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{y}, \mathtt{y} \rightarrow \mathtt{f}\}, \mathtt{f} \rangle$	0.88	0.81	0.80	0.64
$\langle \{\mathtt{b}, \mathtt{b} \rightarrow \mathtt{y}\}, \mathtt{y} \rangle$	0.08	0.24	−0.80	−0.33
$\langle \{\mathtt{y}, \mathtt{y} \rightarrow \mathtt{f}\}, \mathtt{f} \rangle$	0.23	0.07	−0.33	−0.82

In the same way that we consider the strength of an argument, we can consider the strength of a defeasible rule.

Definition 14. The **strength** of rule $\psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi$ is $S^x_P(A)$ where $A = \langle \{\psi_1, \ldots, \psi_n, \psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi\}, \phi \rangle$ and P is a probability distribution and $x \in \{1, 2, 3, 4\}$.

The following result says that if we want compositionally strong arguments, then we only need to consider strong defeasible rules.

Proposition 32. *If argument A is compositionally strong w.r.t. strength measure S^x_P, where $x \in \{1, 2, 3, 4\}$, and threshold $\tau \in [1, -1]$, then the strength of any rule $\rho \in \mathsf{Support}(A)$ is greater than or equal to τ.*

The above result means that if we do select strong defeasible rules for our knowledgebase, then we do not risk missing strong arguments. In other words, by rejecting weak defeasible rules, rules that are not going to lead to strong arguments are eliminated from the knowledgebase.

11. Updating distributions

Our proposal for measuring the strength of an argument depends on the chosen probability distribution, and so it is important to consider that when we are in a specific situation, we may wish to update the distribution. For example, we may have a low belief that a particular bird is a penguin, but when we are in a zoo, we might want to revise the distribution to reflect this since the proportion of birds that are penguins is higher in a zoo. This is a form of updating or commitment, and it can impact the strength of an argument. We address this issue in this section and thereby contribute to our solution for probabilistic argumentation (i.e. Goal 1).

There are various options for updating a probability distribution with new information. In Bayesian updating, the revised belief in a hypothesis H, is $P^*(H) = P(H|E)$ where E is the evidence, and P^* is the updated belief. A shortcoming of this for some applications is that it is assumed the evidence is believed (i.e. $P^*(E) = 1$). Yet often we might not be sure of the belief in the evidence. For instance, if we are in a zoo, we may have a probability distribution such that $P(\mathtt{b}) = 0.1$ for $\mathtt{b}$ denoting bird. Then we may see an animal that we think is a bird but we might have some doubt (uncertainty) and give $P^*(\mathtt{b}) = 0.8$.

To address this need, we can use Jeffrey's rule of conditioning. This can be considered as a way to revise a probability distribution by another probability function [36]. It offers a generalization of Bayes' rule of conditioning. So the updating information does not assume $P^*(E) = 1$, but rather it can be any value in the unit interval. As we have seen so far, we represent evidence by propositions. For Jeffery's rule, we require a set of propositions $\{\beta_1, \ldots, \beta_n\} \subseteq \mathcal{B}(\mathcal{A})$ that are disjoint (i.e. for each i, j, $\{\beta_i, \beta_j\} \vdash \bot$).

Definition 15. For a set of disjoint propositions $\{\beta_1, \ldots, \beta_n\}$, **Jeffrey's rule of conditioning** is $P^*(\alpha) = \sum_i P(\alpha|\beta_i) \times P^*(\beta_i)$.

If we have a single piece of evidence β that is true (i.e. $P^*(\beta) = 1$), then Jeffrey's rule is equivalent to Bayes' rule (i.e. $P^*(\alpha) = P(\alpha|\beta) \times P^*(\beta) = P(\alpha|\beta)$).

Example 28. For the probability distribution P in Example 21, with signature $\langle \mathtt{b}, \mathtt{p}, \mathtt{f} \rangle$, if the update is $P^*(\mathtt{p}) = 0.8$, then the revised probability distribution using Jeffrey's rule of condition is as follows (for each model m, $P^*(m) = P(m|\mathtt{p}) \times P^*(\mathtt{p}) + P(m|\neg\mathtt{p}) \times P^*(\neg\mathtt{p})$). Note, here we treat each model as a conjunction of literals. So for instance for the signature $\langle \mathtt{b}, \mathtt{p}, \mathtt{f} \rangle$, the model 010 is represented by the conjunction $\neg\mathtt{b} \wedge \mathtt{p} \wedge \neg\mathtt{f}$.

	110	101	100	000
P	0.01	0.75	0.04	0.20
P^*	0.80	0.15	0.01	0.04

Consider $A_1 = \langle \mathtt{b}, \mathtt{b} \rightarrow \mathtt{f} \}, \mathtt{f} \rangle$, and $A_2 = \langle \mathtt{p}, \mathtt{p} \rightarrow \neg\mathtt{f} \}, \neg\mathtt{f} \rangle$ with scores given in Table 2. In comparison with Example 21, belief in A_1 falls and A_2 rises, and by the strength measures, A_1 is a much weaker argument, whereas by A_2 is stronger than before.

Table 2
Scores for argument strength in Example 28 where $P(A_i)$ is the probability of acceptability of argument A_i and $P(\mathsf{C}(A_i)|\mathsf{E}(A_i))$ is the probability of the claim of A_i given the evidence in A_i.

	$P(A_i)$	$P(\mathsf{C}(A_i)\|\mathsf{E}(A_i))$	$S^1_{P^*}$	$S^2_{P^*}$	$S^3_{P^*}$	$S^4_{P^*}$
A_1	0.15	0.16	0.16	0.05	−0.69	1.00
A_2	0.80	1.00	0.75	0.95	1.00	0.88

The following result shows that in the special case that we use Bayes' rule of conditioning for an argument A, where the updating is on the premises of the argument (i.e. we assume that $P(\mathsf{E}(A) = 1)$, then the strength measure collapses to zero or is equivalent to the belief in the claim. The reason is that if we already know the premises are true, then the argument has no weight.

Proposition 33. *For an argument A and a probability distribution P, if P^* is the probability distribution obtained by Bayes' rule (i.e. $P^*(\mathsf{C}(A)) = P(\mathsf{C}(A)|\mathsf{E}(A))$), then*

- $S^1_{P^*}(A) = P^*(\mathsf{C}(A))$
- $S^2_{P^*}(A) = 0$
- $S^3_{P^*}(A) = P^*(\mathsf{C}(A)) - P^*(\neg\mathsf{C}(A))$
- $S^4_{P^*}(A) = 1$

We propose the use of Jeffrey's updating rule as an illustration of how we can define updating to capture updates that are not necessarily categorical. Alternatives include minimal change to a probability distribution using distance-based methods [31] or using maximum entropy [13,43,23].

12. Discussion

In this paper, we have provided the first general framework for probability-based measures of argument strength that takes into account the probabilistic necessity and sufficiency of premises for the claim, and also takes into account the probabilistic necessity of the evidence for the complementary claim, and the probabilistic sufficiency of the complementary evidence for the claim. We provide a set of properties for characterizing probability-based measures of argument strength, and we presented four specific probability-based measures each of which provides insights into different dimensions of the strength of arguments. Our investigation therefore addresses the four goals that we identified in the introduction.

We presented the framework for a variant of defeasible logic with a clear semantics. This is so that we can get a clear understanding of the key concepts. We could have used an existing proposal for argumentation, but then the underlying issues we wanted to explore would be less clear in a more complex framework. Nonetheless, we believe this paper provides insights relevant for other argumentation systems, and so in future work, we will adapt existing proposals for structured argumentation systems (e.g. defeasible logic programming [24], ASPIC+ [41], and ABA [61]) to quantify probabilistic strength.

A key potential application of the framework includes analyzing arguments arising in the **audience scenario** given in Section 1. The strength measures give us different ways to judge the arguments, and decide which we regard as strong arguments. For those arguments that we identify as weak, we may ask the presenter of them to justify them. Possibly, they may provide supporting arguments to back-up the arguments in question, and if we are convinced by those supporting arguments, we have the option to then update our probability distribution. The protocol for this, the criteria for being convinced by the supporting arguments, and the method for updating the probability distribution, are beyond the scope of this paper, but this expanded scenario indicates how being able to analyze the strength of the arguments presented by others is potentially useful as part of an argumentation process.

Another key potential application arises in a **persuasion scenario** (for a review of persuasion see [26]). Assume we have a knowledgebase from which we can construct arguments, and we want to persuade another agent (the persuadee) to belief a specific claim using some of the arguments from the knowledgebase. In the dialogue with the persuadee, we want to select the best arguments to present in order to maximize the likelihood that the persuadee will be persuaded. For this, we can construct a probability distribution that reflects what we think the other agent believes about the world. Then using that probability distribution, we could select the stronger arguments to present. In this way, we may increase the probability that the persuasion dialogue is successful. Note, however, probability-based measures of argument strength are only one aspect of determining whether an argument is persuasive. As discussed by Raab [54], relying on the probability of a claim given its premises can be misleading in determining persuasiveness. Also, as shown in our research on persuasion, other dimensions such as the concerns a persuasdee might have about the topic are also important factors in determining the persuasiveness of arguments (see [28] for coverage of concerns in computational persuasion).

A third key application is an **analytical scenario**. If we have acquired knowledge (perhaps from multiple sources), we may want to analyze the quality of the arguments generated from that knowledge. We can construct multiple probability distributions in order to investigate the arguments. Each probability distribution could reflect a possible modelling of the world, and so the change in strength for specific arguments could be investigated. Robustness could be investigated by

identifying how extreme the modelling would be for arguments to be substantially weakened or strengthened. We leave the framework for undertaking robustness analysis to future work.

In future work, we will also consider generalizing the framework to use alternative base logics including classical logic and paraconsistent logics instead of defeasible logic, and to consider dropping the minimality and/or consistency constraint for arguments. For this we can consider a wider range of characterization properties including the following (where Cn is the consequence closure function for a base logic such as classical logic or a paraconsistent logic) which consider the effect of inconsistency, of non-minimality, and of tautologies, on probabilistic strength.

- If $\mathsf{E}(A) = \emptyset$, then $S(A) = 0$
- If $\mathsf{E}(A) \vdash \bot$, then $S(A) \leq 0$
- If $\mathsf{E}(A) \cup \{\mathsf{C}(A)\} \vdash \bot$, then $S(A) \leq 0$
- If $\mathsf{Cn}(\mathsf{Support}(A)) \subseteq \mathsf{Cn}(\mathsf{Support}(B))$, and $\mathsf{C}(A) = \mathsf{C}(A)$, then $S(A) > S(B)$.

Also, in future work, we will consider generalizing our proposals for argument strength to incorporate another approach for argument strength that defines the strength of an argument as a function of the probability of the premises and the probability of the claim given the premises [46]. If the probability of the claim given the premises is x, and the probability of the premises is y, then $z' = xy$ is a lower bound on the probability of the claim, and $z'' = xy + 1 - y$ is an upper bound on the probability of the claim. From the lower and upper bounds, Pfeifer gives the following definition of the strength of an argument A, denoted $S(A)$, where $(1 - (z'' - z'))$ is the *precision*, and $(z' + z'')/2$ is the *location*.

$$S(A) = (1 - (z'' - z')) \times \frac{(z' + z'')}{2}$$

In the above definition, the precision is higher when the gap between the lower and upper bounds is smaller, and the location is higher when the lower and upper bounds are closer to one. So a stronger argument is one where the belief in the claim has higher precision and higher location. This provides an intuitive and potentially valuable proposal for analyzing probabilistic arguments that could be harnessed in our framework to give further valuable measures.

Finally, in future work, we will consider practical issues for obtaining and efficiently managing probability distributions over models. It is possible to crowd-source beliefs in individual argument (or at least, proxies for belief in individual arguments), as shown in [30,47,29]. These beliefs can then be used as constraints in linear optimization to identify satisfying probability distribution, or by assuming appropriate conditional independences between the atoms in the models [27].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Anthony Hunter reports financial support was provided by The Alan Turing Institute.

Acknowledgements

The research in this paper was partly supported by funding from The Alan Turing Institute for the *Interpretable and Explainable Deep Learning for Natural Language Understanding and Common Sense Reasoning* project. The author is grateful to the anonymous reviewers for valuable suggestions for improving the paper.

Appendix A

Lemma 1. *For a probability distribution P, and formulae $\phi, \psi \in \mathcal{B}(\mathcal{A})$. $P(\phi|\neg\psi)\#P(\phi|\psi)$ iff $P(\phi)\#P(\phi|\psi)$, where $\# \in \{<, \leq, =, \geq, >\}$.*

Proof. $P(\phi|\neg\psi)\#P(\phi|\psi)$ iff $P(\phi|\neg\psi)(1 - P(\psi)) \ \# \ P(\phi|\psi)(1 - P(\psi))$ iff $P(\phi|\neg\psi)(1 - P(\psi)) \ \# \ (P(\phi|\psi) - P(\phi|\psi)P(\psi))$ iff $P(\phi|\neg\psi)P(\neg\psi) \ \# \ (P(\phi|\psi) - P(\phi|\psi)P(\psi))$ iff $(P(\phi|\neg\psi)P(\neg\psi) + P(\phi|\psi)P(\psi)) \ \# \ P(\phi|\psi)$ iff $(P(\phi \wedge \neg\psi) + P(\phi \wedge \psi)) \ \# \ P(\phi|\psi)$ iff $P(\phi)\#P(\phi|\psi)$. □

Lemma 2. *For a probability distribution P, and formulae $\phi, \psi \in \mathcal{B}(\mathcal{A})$. $P(\phi|\psi)\#P(\phi)$ iff $P(\psi|\phi)\#P(\psi)$, where $\# \in \{<, \leq, =, \geq, >\}$.*

Proof. $P(\phi|\psi)\#P(\phi)$ iff $P(\phi \wedge \psi)/P(\psi)\#P(\phi)$ iff $P(\phi \wedge \psi)\#P(\psi)P(\phi)$ iff $P(\phi \wedge \psi)/P(\phi)\#P(\psi)$ iff $P(\psi|\phi)\#P(\psi)$ □

Lemma 3. *For a probability distribution P, and formulae $\phi, \psi \in \mathcal{B}(\mathcal{A})$. $P(\phi|\psi)\#P(\phi|\neg\psi)$ iff $P(\psi|\phi)\#P(\psi|\neg\phi)$ where $\# \in \{<, \leq, =, \geq, >\}$.*

Proof. By Lemma 1, $P(\phi|\psi)\#P(\phi|\neg\psi)$ iff $P(\phi|\psi)\#P(\phi)$. Therefore, by Lemma 2, $P(\phi|\psi)\#P(\phi)$ iff $P(\psi|\phi)\#P(\psi)$. And so by Lemma 1, $P(\psi|\phi)\#P(\psi)$ iff $P(\psi|\phi)\#P(\psi|\neg\phi)$. □

Proposition 1. *For $\Delta \subseteq \mathcal{F}(\mathcal{A})$, $\phi \in \mathcal{L}(\mathcal{A})$, $\Delta \vdash \phi$ iff $\Delta \models \phi$.*

Proof. (Assume Δ is consistent) This result can be shown via the notion of a proof tree where ϕ is at the root, each leaf is a literal in Δ, and each non-leaf node ϕ' is such that there is a rule $\psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi' \in \Delta$ and each $\psi_i \in \{\psi_1, \ldots, \psi_n\}$ is a child of ϕ'. So $\Delta \vdash \phi$ holds iff there is such a proof tree where each branch is finite. This is sound because for each leaf ψ, $\psi \in \Delta$, and therefore $\psi \in \mathsf{Infer}(\Delta)$. Hence, for all models $m \in \mathsf{Models}(\Delta)$, $m \in \mathsf{Sat}(\psi)$, and so $\Delta \models \psi$, and because for each non-leaf ψ', there is a $\rho = \psi_1 \wedge \ldots \wedge \psi_n \rightarrow \psi' \in \Delta$ and $i \in \mathbb{N}$, such that $\mathsf{Inf}^{i+1}(\Delta) = \mathsf{Inf}^{i}(\Delta) \cup \{\mathsf{Head}(\rho) \mid \text{for all } \phi \in \mathsf{Tail}(\rho), \phi \in \mathsf{Inf}^{i}(\Delta)\}$. Therefore, for all $\phi \in \mathsf{Tail}(\rho), \phi \in \mathsf{Infer}(\Delta)$ and $\psi' \in \mathsf{Infer}(\Delta)$. Hence, for all models $m \in \mathsf{Models}(\Delta)$, $\phi \in \mathsf{Tail}(\rho)$, $m \in \mathsf{Sat}(\phi)$ and $m \in \mathsf{Sat}(\psi')$. Hence, $\phi \in \mathsf{Tail}(\rho)$, $\Delta \models \phi$ and $\Delta \models \psi'$. Furthermore, the use of proof trees is complete since for every ϕ such that $\Delta \models \phi$, there is a proof tree where each branch is finite. Therefore, $\Delta \vdash \phi$ iff $\Delta \models \phi$. (Assume Δ is inconsistent) So the consequence relation entails any literal because of proof rule 3 (ex falso quadlibet). Therefore for any β, $\Delta \vdash \beta$. Also the entailment relation for defeasible logic is trivializable in the sense that any literal is an inference from inconsistency. Therefore for any β, $\Delta \models \beta$. This is because when Δ is inconsistent, $\mathsf{Models}(\Delta) = \emptyset$, and therefore for any $\beta \in \mathcal{L}$, $\mathsf{Models}(\Delta) \subseteq \mathsf{Models}(\beta)$. Therefore, $\Delta \vdash \phi$ iff $\Delta \models \phi$. □

Proposition 2. *For probability distribution P, if B attacks A, then $P(A) + P(B) \leq 1$.*

Proof. The definition of rebuttal implies $P(\mathsf{C}(A)) + P(\mathsf{C}(B)) \leq 1$ and the definition of direct undercut implies $P(\bigwedge \mathsf{Facts}(A)) + P(\mathsf{C}(B)) \leq 1$. In either case, $P(\bigwedge(\mathsf{Facts}(A) \cup \{\mathsf{C}(A)\})) + P(\bigwedge(\mathsf{Facts}(B) \cup \{\mathsf{C}(B)\})) \leq 1$. Therefore, $P(A) + P(B) \leq 1$. □

Proposition 3. *For a probability distribution P, premise ψ, and claim ϕ, if P is uniform, then $P(\phi|\psi) = P(\psi|\phi) = P(\phi|\neg\psi) = P(\psi|\neg\phi) = 0.5$.*

Proof. Let $P(\phi \wedge \psi) = 0.25$, $P(\phi \wedge \neg\psi) = 0.25$, $P(\neg\phi \wedge \psi) = 0.25$, and $P(\neg\phi \wedge \neg\psi) = 0.25$. So P is the uniform distribution. So $P(\phi|\psi) = P(\phi \wedge \psi)/P(\psi) = 0.25/0.5 = 0.5$. We obtain the same ratios for $P(\psi|\phi)$, $P(\phi|\neg\psi)$, and $P(\psi|\neg\phi)$. □

Proposition 4. *For a probability distribution P, premise ψ and claim ϕ, $P(\phi|\psi) = 1$ and $P(\phi|\neg\psi) = 1$ iff $P(\phi \wedge \psi) + P(\phi \wedge \neg\psi) = 1$ and $P(\phi \wedge \psi) > 0$ and $P(\phi \wedge \neg\psi) > 0$.*

Proof. $P(\phi|\psi) = 1$ and $P(\phi|\neg\psi) = 1$ iff $P(\phi \wedge \psi)/P(\psi) = 1$ and $P(\phi|\neg\psi)/P(\neg\psi) = 1$ iff $P(\phi \wedge \psi) \neq 0$ and $P(\neg\phi \wedge \psi) = 0$ and $P(\phi \wedge \neg\psi) \neq 0$ and $P(\neg\phi \wedge \neg\psi) = 0$ iff $P(\phi \wedge \psi) + P(\phi \wedge \neg\psi) = 1$ and $P(\phi \wedge \psi) > 0$ and $P(\phi \wedge \neg\psi) > 0$. □

Proposition 5. *For a probability distribution P, premise ϕ and claim ψ, $P(\phi|\psi) = 1$ and $P(\psi|\phi) = 1$ iff $P(\phi \wedge \psi) = P(\phi)$ and $P(\phi \wedge \psi) = P(\psi)$.*

Proof. $P(\psi|\phi) = 1$ and $P(\phi|\alpha) = 1$ iff $P(\psi \wedge \phi)/P(\phi) = 1$ and $P(\phi \wedge \psi)/P(\psi) = 1$ iff $P(\phi \wedge \psi) = P(\phi)$ and $P(\phi \wedge \psi) = P(\psi)$. □

Proposition 6. *Let P be a probability distribution, and $\phi, \psi, \psi_1, \psi_2 \in \mathcal{B}(\mathcal{A})$. (1) $P(\phi|\psi) = 1$ iff $\{\psi\} \vdash \phi$; (2) $P(\phi|\psi) = 0$ iff $\{\psi\} \vdash \neg\phi$; And (3) $P(\phi_1|\psi) \leq P(\phi_2|\psi)$ if $\{\phi_1\} \vdash \phi_2$.*

Proof. (1) $P(\phi|\psi) = 1$ iff $P(\phi \wedge \psi)/P(\psi) = 1$ iff $P(\phi \wedge \psi) = P(\psi)$ iff $\mathsf{Sat}(\psi) = \mathsf{Sat}(\phi \wedge \psi)$ iff $\{\psi\} \vdash \phi$. (2) $P(\phi|\psi) = 0$ iff $P(\phi \wedge \psi)/P(\psi) = 0$ iff $\mathsf{Sat}(\phi \wedge \psi) = \emptyset$ iff $\{\psi\} \vdash \neg\phi$. (3) Assume $\{\phi_1\} \vdash \phi_2$. So $\mathsf{Sat}(\phi_1 \wedge \psi) \subseteq \mathsf{Sat}(\phi_2 \wedge \psi)$. So $P(\phi_1 \wedge \psi) \leq P(\phi_2 \wedge \psi)$. So $P(\phi_1 \wedge \psi)/P(\psi) \leq P(\phi_2 \wedge \psi)/P(\psi)$. So $P(\phi_1|\psi) \leq P(\phi_2|\psi)$. □

Proposition 7. *For any argument A, and probability distribution P,*

1. $S^1_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$
2. $S^1_P(A) = -1$ *iff* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$
3. $S^2_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$
4. $S^2_P(A) = -1$ *iff* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$
5. $S^3_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$
6. $S^3_P(A) = -1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$
7. $S^4_P(A) = 1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$
8. $S^4_P(A) = -1$ *iff* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ *and* $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) \neq 0$

Proof. We consider each case as follows. (1) $S^1_P(A) = 1$ iff $(P(\mathsf{C}(A)|\mathsf{E}(A)) = 1$ and $P(\mathsf{C}(A)|\neg\mathsf{E}(A))$ is zero or undefined) iff $((P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = P(\mathsf{C}(A) \wedge \mathsf{E}(A)) + P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)))$ and $((P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0))$ iff $(P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$

and $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$ and $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$). (2) $S^1_P(A) = -1$ iff ($P(\mathsf{C}(A)|\mathsf{E}(A))$ is zero or undefined and $P(\mathsf{C}(A)|\neg\mathsf{E}(A)) = 1$) iff (($P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$) and (($P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) + P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A))$))) iff ($P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) \neq 0$ and ($P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ and $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$)). (3) $S^2_P(A) = 1$ iff ($P(\mathsf{E}(A)|\mathsf{C}(A)) = 1$ and $P(\mathsf{E}(A)|\neg\mathsf{C}(A))$ is zero or undefined) iff (($P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = P(\mathsf{C}(A) \wedge \mathsf{E}(A)) + P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A))$) and (($P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$)) iff $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ and $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$ and $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$. (4) $S^2_P(A) = -1$ iff ($P(\mathsf{E}(A)|\mathsf{C}(A))$ is zero or undefined and $P(\neg\mathsf{E}(A)|\mathsf{C}(A)) = 1$) iff (($P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$) and (($P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) + P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A))$))) iff $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ and $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ and $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$. (5) $S^3_P(A) = 1$ iff ($P(\mathsf{C}(A)|\mathsf{E}(A)) = 1$ and $P(\neg\mathsf{C}(A)|\mathsf{E}(A))$ is zero or undefined) iff $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ and $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$. (6) $S^3_P(A) = -1$ iff ($P(\mathsf{C}(A)|\mathsf{E}(A))$ is zero or undefined and $P(\neg\mathsf{C}(A)|\mathsf{E}(A)) = 1$) iff $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ and $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$. (7) $S^4_P(A) = 1$ iff ($P(\mathsf{E}(A)|\mathsf{C}(A)) = 1$ and $P(\neg\mathsf{E}(A)|\mathsf{C}(A))$ is zero or undefined) iff $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) \neq 0$ and $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$. (8) $S^4_P(A) = -1$ iff ($P(\mathsf{E}(A)|\mathsf{C}(A))$ is zero or undefined and $P(\neg\mathsf{E}(A)|\mathsf{C}(A)) = 1$) iff $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ and $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) \neq 0$. □

Corollary 1. *For any argument A, and probability distribution P,*

1. $S^1_P(A) = 1$ *iff* $S^2_P(A) = 1$
2. $S^3_P(A) = 1$ *if* $S^1_P(A) = 1$ *or* $S^2_P(A) = 1$
3. $S^3_P(A) = -1$ *if* $S^2_P(A) = -1$
4. $S^4_P(A) = 1$ *if* $S^1_P(A) = 1$ *or* $S^2_P(A) = 1$
5. $S^4_P(A) = -1$ *if* $S^1_P(A) = -1$

Proof. Results follow directly from Proposition 7. □

Proposition 8. *For measures S^1_P, S^2_P, S^3_P, and S^4_P, S^1_P is not independent of S^2_P S^1_P is independent of S^3_P, S^1_P is independent of S^4_P, S^2_P is independent of S^3_P, S^2_P is independent of S^4_P, and S^3_P is independent of S^4_P.*

Proof. Let A be an argument, and P be a probability distribution where $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = x_0$, $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = x_1$, $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = x_2$, and $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = x_3$. Also let $\# \in \{<, \leq, =, \geq, >\}$ and $v \in [-1, 1]$. We show (1) holds and give counterexamples for (2) to (6). (1) $S^1_P(A)\#v$ iff $x_0/(x_0+x_2)\#x_1/(x_1+x_3)$ iff $x_0(x_1+x_3)\#x_1(x_0+x_2)$ iff $x_0x_1+x_0x_3\#x_0x_1+x_1x_2$ iff $x_0x_3\#x_1x_2$ iff $x_0x_2 + x_0x_3\#x_0x_2 + x_1x_2$ iff $x_0(x_2 + x_3)\#x_2(x_0 + x_1)$ iff $x_0/(x_0 + x_1)\#x_2/(x_2 + x_3)$ iff $S^2_P(A)\#v$. So S^1_P is not independent of S^2_P. (2) Let $x_0 = 0.1$, $x_1 = 0.1$, $x_2 = 0.2$, and $x_3 = 0.6$. So $P(\mathsf{C}(A) \mid \mathsf{E}(A)) = 0.1/0.3$, and $P(\mathsf{C}(A) \mid \neg\mathsf{E}(A)) = 0.1/0.7$. So $S^1_P(A) = 1/3 - 1/7 = 4/21$ and $S^3_P(A) = 1/3 - 2/3 = -1/3$. So $S^1_P(A) > 0$ and $S^3_P(A) < 0$. So S^1_P is independent of S^3_P. (3) Let $x_0 = 0.1$, $x_1 = 0.2$, $x_2 = 0.2$, and $x_3 = 0.5$. So $P(\mathsf{C}(A) \mid \mathsf{E}(A)) = 0.1/0.3$, $P(\mathsf{C}(A) \mid \neg\mathsf{E}(A)) = 0.2/0.7$, and $P(\mathsf{E}(A) \mid \mathsf{C}(A)) = 0.1/0.3$. So $S^1_P(A) = 1/3 - 2/7 = 1/21$ and $S^4_P(A) = 1/3 - 2/3 = -1/3$. So $S^1_P(A) > 0$ and $S^4_P(A) < 0$. So S^1_P is independent of S^4_P. (4) Let $x_0 = 0.1$, $x_1 = 0.1$, $x_2 = 0.2$, and $x_3 = 0.6$. So $P(\mathsf{C}(A) \mid \mathsf{E}(A)) = 0.1/0.3$, $P(\mathsf{E}(A) \mid \mathsf{C}(A)) = 0.1/0.2$, and $P(\mathsf{E}(A) \mid \neg\mathsf{C}(A)) = 0.2/0.8$. So $S^2_P(A) = 1/2 - 1/4 = 1/4$ and $S^3_P(A) = -1/3$. So $S^2_P(A) > 0$ and $S^3_P(A) < 0$. So S^2_P is independent of S^3_P. (5) Let $x_0 = 0.1$, $x_1 = 0.2$, $x_2 = 0.1$, and $x_3 = 0.6$. So $P(\mathsf{E}(A) \mid \mathsf{C}(A)) = 0.1/0.3$, and $P(\mathsf{E}(A) \mid \neg\mathsf{C}(A)) = 0.1/0.7$. So $S^2_P(A) = 1/3 - 1/7 = 4/21$ and $S^4_P(A) = -1/3$. So $S^2_P(A) > 0$ and $S^4_P(A) < 0$. So S^2_P is independent of S^4_P. (6) Let $x_0 = 0.2$, $x_1 = 0.3$, $x_2 = 0.1$, and $x_3 = 0.4$. So $P(\mathsf{C}(A) \mid \mathsf{E}(A)) = 0.2/0.3$, and $P(\mathsf{E}(A) \mid \mathsf{C}(A)) = 0.2/0.5$. So $S^3_P(A) = 1/3$ and $S^4_P(A) = -1/5$. So $S^3_P(A) > 0$ and $S^4_P(A) < 0$. So S^3_P is independent of S^4_P. □

Proposition 9. *For any argument A, for any $\delta \in (0, 1]$, there is a non-zero probability distribution P such that $S^i_P(A) = 1 - \delta$ for all $i \in \{1, 2, 3, 4\}$.*

Proof. For $x \in [0, 1]$, let $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = (0.5 - x)$, $P(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = x$, $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = x$, and $P(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = (0.5 - x)$. Hence, $S^1_P(A) = S^2_P(A) = S^3_P(A) = S^4_P(A)$. From P, we have $S^1_P(A) = ((0.5-x)/0.5) - (x/0.5)) = ((0.5-x)-x)/0.5 = 1 - (2x/0.5) = 1 - 4x$. So if we let $x = \delta/4$, and let $1 - \delta = 1 - 4x$, then we have $S^i_P(A) = 1 - \delta$ for all $i \in \{1, 2, 3, 4\}$. □

Proposition 10. [EQ1] *If $\mathsf{E}(A) \equiv \mathsf{E}(B)$ and $\mathsf{C}(A) \equiv \mathsf{C}(B)$, then $S(A) = S(B)$. Holds for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Direct from definition. □

Proposition 11. [EQ2] *If $\mathsf{E}(A) \equiv \neg\mathsf{C}(B)$ and $\mathsf{C}(A) \equiv \neg\mathsf{E}(B)$, then $S(A) = S(B)$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$ and let B be an argument, where $\mathsf{C}(B) = \neg\psi$ and $\mathsf{E}(B) = \neg\phi$. Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.2$, $P(\neg\phi \wedge \psi) = 0.3$, and $P(\neg\phi \wedge \neg\psi) = 0.4$. For A, $P(\phi \mid \psi) = 1/4$, $P(\phi \mid \neg\psi) = 2/6$, $P(\psi \mid \phi) = 1/3$, and $P(\psi \mid \neg\phi) = 3/7$. For B, $P(\neg\psi \mid \neg\phi) = 4/7$, $P(\neg\psi \mid \phi) = 2/3$, $P(\neg\phi \mid \neg\psi) = 4/6$, and $P(\neg\phi \mid \psi) = 3/4$. ($S^1_P$) So $S^1_P(A) = 1/4 - 2/6 = -1/12$ and $S^1_P(B) = 4/7 - 2/3 = -2/21$. So $S^1_P(A) \neq S^1_P(B)$. (S^2_P) So $S^2_P(A) = 1/3 - 3/7 = -2/21$

and $S^2_P(B) = 4/6 - 3/4 = -1/12$. So $S^2_P(A) \neq S^2_P(B)$. (S^3_P) So $S^3_P(A) = 1/4 - 3/4 = -1/2$ and $S^3_P(B) = 4/7 - 3/7 = 1/7$. So $S^3_P(A) \neq S^3_P(B)$. (S^4_P) So $S^4_P(A) = 1/3 - 2/3 = -1/3$ and $S^4_P(B) = 4/6 - 2/6 = 1/3$. So $S^4_P(A) \neq S^4_P(B)$. □

Proposition 12. [MAX1] *If* $\mathsf{E}(A) \equiv \mathsf{C}(A)$, *then* $S(A) = 1$. *Holds for* S^1_P *and* S^3_P, *and fails for* S^2_P *and* S^4_P

Proof. Assume $\mathsf{E}(A) \equiv \mathsf{C}(A)$. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P) $P(\mathsf{C}(A)|\mathsf{E}(A)) = 1$ and $P(\mathsf{C}(A)|\neg\mathsf{E}(A))$ is undefined. So $S^1_P(A) = 1$. (S^2_P) Let $P(\phi \wedge \psi) = P(\phi \wedge \neg\psi)$ and $P(\neg\phi \wedge \psi) = P(\neg\phi \wedge \neg\psi) = 0$. So $P(\psi|\phi) = 1/2$ and $P(\psi|\neg\phi)$ is undefined. So $S^2_P(A) = 1/2$. (S^3_P) $P(\mathsf{C}(A)|\mathsf{E}(A)) = 1$ and $P(\neg\mathsf{C}(A)|\mathsf{E}(A)) = 0$. So $S^3_P(A) = 1$. (S^4_P) Using the same example as for S^2_P, $P(\psi|\phi) = 1/2$ and $P(\neg\psi|\phi) = 1/2$. So $S^4_P(A) = 0$. □

Proposition 13. [MAX2] *If* $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 1$, *then* $S(A) = 1$. *Holds for* S^1_P, S^2_P, S^3_P, *and* S^4_P.

Proof. Assume $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 1$. So $P(\mathsf{C}(A)|\mathsf{E}(A)) = 1$, $P(\mathsf{E}(A)|\mathsf{C}(A)) = 1$, $P(\mathsf{C}(A)|\neg\mathsf{E}(A))$ is undefined, and $P(\mathsf{E}(A)|\neg\mathsf{C}(A))$ is undefined. Hence, $S^1_P(A) = S^2_P(A) = S^3_P(A) = S^4_P(A) = 1$. □

Proposition 14. [ZERO1] *If* P *is a uniform probability distribution, then* $S(A) = 0$. *Holds for* S^1_P, S^2_P, S^3_P, *and* S^4_P.

Proof. Assume P is uniform. (S^1_P) $P(\mathsf{C}(A)|\mathsf{E}(A)) = 0.5$ and $P(\mathsf{C}(A)|\neg\mathsf{E}(A)) = 0.5$. So $S^1_P(A) = 0$. (S^2_P) $P(\mathsf{E}(A)|\mathsf{C}(A)) = 0.5$ and $P(\mathsf{E}(A)|\neg\mathsf{C}(A)) = 0.5$. So $S^2_P(A) = 0$. (S^3_P) $P(\mathsf{C}(A)|\mathsf{E}(A)) = 0.5$ and $P(\neg\mathsf{C}(A)|\mathsf{E}(A)) = 0.5$. So $S^3_P(A) = 0$. (S^4_P) $P(\mathsf{E}(A)|\mathsf{C}(A)) = 0.5$ and $P(\neg\mathsf{E}(A)|\mathsf{C}(A)) = 0.5$. So $S^4_P(A) = 0$. □

Proposition 15. [ZERO2] *If* $P(\mathsf{C}(A)|\mathsf{E}(A)) = P(\mathsf{C}(A))$, *then* $S(A) = 0$. *Holds for* S^1_P *and* S^2_P, *and fails for* S^3_P *and* S^4_P.

Proof. Assume $P(\mathsf{C}(A)|\mathsf{E}(A)) = P(\mathsf{C}(A))$. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P) By Lemma 1, $P(\mathsf{C}(A)) = P(\mathsf{C}(A)|\mathsf{E}(A))$ iff $P(\mathsf{C}(A)|\neg\mathsf{E}(A)) = P(\mathsf{C}(A)|\mathsf{E}(A))$. So $S(A) = 0$. (S^2_P) By Lemma 1, $P(\mathsf{C}(A)) = P(\mathsf{C}(A)|\mathsf{E}(A))$ iff $P(\mathsf{C}(A)|\neg\mathsf{E}(A)) = P(\mathsf{C}(A)|\mathsf{E}(A))$. Then from Lemma 3, $P(\mathsf{C}(A)|\mathsf{E}(A)) = P(\mathsf{C}(A)|\neg\mathsf{E}(A))$ iff $P(\mathsf{E}(A)|\mathsf{C}(A)) = P(\mathsf{E}(A)|\neg\mathsf{C}(A))$. So $S^2_P(A) = 0$. (S^3_P) Let $P(\phi \wedge \psi) = 0.4$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi) = 4/5$, $P(\phi \mid \psi) = 4/5$ and $P(\neg\phi \mid \psi) = 1/5$. So $P(\phi \mid \psi) = P(\phi)$, but $S^3_P(A) = 3/5$. (S^4_P) Let $P(\phi \wedge \psi) = 0.4$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.4$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi) = 1/2$, $P(\phi \mid \psi) = 1/2$, $P(\psi \mid \phi) = 4/5$, and $P(\neg\psi \mid \phi) = 1/5$. So $P(\phi \mid \psi) = P(\phi)$, but $S^4_P(A) = 3/5$. □

Proposition 16. [ZERO3] *If* $P(\mathsf{E}(A)|\mathsf{C}(A)) = P(\mathsf{E}(A)|\neg\mathsf{C}(A))$, *then* $S(A) = 0$. *Holds for* S^1_P *and* S^2_P, *and fails for* S^3_P *and* S^4_P.

Proof. (S^1_P) Follows via Lemma 3. (S^2_P) Follows directly from definition. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^3_P) Let $P(\phi \wedge \psi) = 0.4$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\psi \mid \phi) = P(\psi \mid \neg\phi) = 1/2$, $P(\phi \mid \psi) = 4/5$, and $P(\phi \mid \neg\psi) = 4/5$ but $S^3_P(A) = 3/5$. (S^4_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.4$. So $P(\psi \mid \phi) = P(\psi \mid \neg\phi) = 1/5$, and $P(\neg\psi \mid \phi) = 4/5$, but $S^4_P(A) = -3/5$ □

Proposition 17. [ZERO4] *If* $P(\mathsf{C}(A)|\mathsf{E}(A)) = P(\neg\mathsf{C}(A)|\mathsf{E}(A))$, *then* $S(A) = 0$. *Holds for* S^3_P, *and fails for* S^1_P, S^2_P, *and* S^4_P.

Proof. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P and S^2_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.7$. So $P(\phi|\psi) = 1/2$, $P(\phi|\neg\psi) = 1/8$, $P(\psi|\phi) = 1/2$, and $P(\psi|\neg\phi) = 1/8$. So $P(\phi|\psi) = P(\neg\phi|\psi)$, but $S^1_P(A) = 3/8$ and $S^2_P(A) = 3/8$. (S^3_P) Follows directly from definition. (S^4_P) Let $P(\phi \wedge \psi) = 0.4$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.4$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi|\psi) = 1/2$, $P(\neg\phi|\psi) = 1/2$, $P(\psi|\phi) = 4/5$, and $P(\neg\psi|\phi) = 1/5$. So $P(\phi|\psi) = P(\neg\phi|\psi)$, but $S^4_P(A) = 3/5$. □

Proposition 18. [ZERO5] *If* $P(\mathsf{E}(A)|\mathsf{C}(A)) = P(\neg\mathsf{E}(A)|\mathsf{C}(A))$, *then* $S(A) = 0$. *Holds for* S^4_P, *and fails for* S^1_P, S^2_P, *and* S^3_P.

Proof. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P and S^2_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.7$. So $P(\phi|\psi) = 1/2$, $P(\phi|\neg\psi) = 1/8$, $P(\psi|\phi) = 1/2$, and $P(\psi|\neg\phi) = 1/8$. So $P(\psi|\phi) = P(\neg\psi|\phi)$, but $S^1_P(A) = 3/8$ and $S^2_P(A) = 3/8$. (S^3_P) Let $P(\phi \wedge \psi) = 0.4$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi|\psi) = 4/5$, $P(\phi|\neg\psi) = 4/5$, $P(\psi|\phi) = 1/2$, and $P(\psi|\neg\phi) = 1/2$. So $P(\psi|\phi) = P(\neg\psi|\phi)$, but $S^3_P(A) = 3/5$. (S^4_P) Follows directly from definition. □

Proposition 19. [POS1] *If* $S(A) > 0$, *then* $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0$. *Holds for* S^1_P, S^2_P, S^3_P, *and* S^4_P.

Proof. (S^1_P) Assume $S^1_P(A) > 0$. So $P(\mathsf{C}(A))|\mathsf{E}(A)) > P(\mathsf{C}(A))|\neg\mathsf{E}(A))$. So $P(\mathsf{C}(A))|\mathsf{E}(A)) > 0$. ($S^2_P$) Assume $S^2_P(A) > 0$. So $P(\mathsf{E}(A))|\mathsf{C}(A)) > P(\mathsf{E}(A))|\neg\mathsf{C}(A))$. So $P(\mathsf{E}(A))|\mathsf{C}(A)) > 0$. So $P(\mathsf{C}(A)) \wedge \mathsf{E}(A)) > 0$ and $P(\mathsf{E}(A)) > 0$. So $P(\mathsf{C}(A))|\mathsf{E}(A)) > 0$. ($S^3_P$) Assume $S^3_P(A) > 0$. So $P(\mathsf{C}(A))|\mathsf{E}(A)) > P(\neg\mathsf{C}(A))|\mathsf{E}(A))$. So $P(\mathsf{C}(A))|\mathsf{E}(A)) > 0$. ($S^4_P$) $S^4_P(A) > 0$. So $P(\mathsf{E}(A))|\mathsf{C}(A)) > P(\neg\mathsf{E}(A))|\mathsf{C}(A))$. So $P(\mathsf{E}(A))|\mathsf{C}(A)) > 0$. So $P(\mathsf{C}(A)) \wedge \mathsf{E}(A)) > 0$ and $P(\mathsf{E}(A)) > 0$. So $P(\mathsf{C}(A))|\mathsf{E}(A)) > 0$. □

Proposition 20. [POS2] *If $S(A) > 0$, then $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0.5$. Holds for S^3_P, and fails for S^1_P, S^2_P, and S^4_P.*

Proof. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P) Let $P(\phi \wedge \psi) = 0.2$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.3$, and $P(\neg\phi \wedge \neg\psi) = 0.4$. So $S^1_P(A) > 0$ (as $P(\phi|\psi) = 2/5$ and $P(\phi|\neg\psi) = 1/5$), hence $S^1_P(A) = 1/5$), but $P(\phi|\psi) < 0.5$. (S^2_P) Let $P(\phi \wedge \psi) = 0.19$, $P(\phi \wedge \neg\psi) = 0.01$, $P(\neg\phi \wedge \psi) = 0.2$, and $P(\neg\phi \wedge \neg\psi) = 0.6$. So $S^2_P(A) > 0$ (as $P(\psi|\phi) = 0.19/0.2$ and $P(\psi|\neg\phi) = 0.2/0.8$), hence $S^2_P(A) = 0.7$), but $P(\phi|\psi) < 0.5$ (as $P(\phi|\psi) = 0.19/0.39$). (S^3_P) Direct from definition. (S^4_P) Let $P(\phi \wedge \psi) = 0.2$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.6$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi|\psi) = 1/4$, $P(\psi|\phi) = 2/3$, and $P(\neg\psi|\phi) = 1/3$. So $S^4_P(A) > 0$, but $P(\phi|\psi) < 0.5$. □

Proposition 21. [POS3] *If $S(A) > 0$, then $P(\mathsf{E}(A)|\mathsf{C}(A)) > 0.5$. Holds for S^4_P, and fails for S^1_P, S^2_P, and S^3_P.*

Proof. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P) Let $P(\phi \wedge \psi) = 0.19$, $P(\phi \wedge \neg\psi) = 0.2$, $P(\neg\phi \wedge \psi) = 0.01$, and $P(\neg\phi \wedge \neg\psi) = 0.6$, and so $S^1_P(A) > 0$ (as $P(\phi|\psi) = 0.19/0.2$) and $P(\phi|\neg\psi) = 0.2/0.8$), hence $S^1_P(A) = 0.7$, but $P(\psi|\phi) < 0.5$ (as $P(\psi|\phi) = 0.19/0.39$). (S^2_P) Let $P(\phi \wedge \psi) = 0.29$, $P(\phi \wedge \neg\psi) = 0.3$, $P(\neg\phi \wedge \psi) = 0.01$, and $P(\neg\phi \wedge \neg\psi) = 0.4$, and so $P(\psi|\phi) = 0.29/0.59$ and $P(\psi|\neg\phi) = 1/50$). So $S^2_P(A) = 0.47$. So $S^2_P(A) > 0$, but $P(\psi|\phi) < 0.5$. (S^3_P) Let $P(\phi \wedge \psi) = 0.2$, $P(\phi \wedge \neg\psi) = 0.6$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi \mid \psi) = 2/3$ and $P(\psi \mid \phi) = 1/4$. So $S^3_P(A) > 0$, but $P(\psi|\phi) < 0.5$. (S^4_P) Direct from definition. □

Proposition 22. [POS4] *If $P(\mathsf{C}(A)|\mathsf{E}(A)) > P(\mathsf{C}(A)$, then $S(A) > 0$. Holds for S^1_P, and fails for S^2_P, S^3_P, and S^4_P.*

Proof. Assume $P(\mathsf{C}(A)|\mathsf{E}(A)) > P(\mathsf{C}(A)$. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P) By Lemma 1, $P(\mathsf{C}(A)|\mathsf{E}(A)) > P(\mathsf{C}(A)|\neg\mathsf{E}(A))$. So $P(\mathsf{C}(A)|\mathsf{E}(A)) - P(\mathsf{C}(A)|\neg\mathsf{E}(A)) > 0$. So $S^1_P(A) > 0$. (S^2_P) Let $P(\phi \wedge \psi) = 0.2$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.3$, and $P(\neg\phi \wedge \neg\psi) = 0.4$. So $P(\phi|\psi) = 2/5$, $P(\phi) = 3/10$, $P(\psi|\phi) = 2/3$, $P(\psi|\neg\phi) = 3/4$, so $S^2_P(A) = 2/3 - 3/4 = -1/12$, and so $S^2_P(A) < 0$. (S^3_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.1$, $P(\neg\phi \wedge \psi) = 0.3$, and $P(\neg\phi \wedge \neg\psi) = 0.4$. So $P(\phi|\psi) = 1/4$, $P(\phi) = 2/10$, $P(\neg\phi|\psi) = 3/4$, so $S^3_P(A) = 1/4 - 3/4 = -1/2$, and so $S^3_P(A) < 0$. (S^4_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.2$, $P(\neg\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \neg\psi) = 0.4$. So $P(\phi|\psi) = 1/2$, $P(\phi) = 3/10$, $P(\psi|\phi) = 1/3$, $P(\neg\psi|\phi) = 2/3$, so $S^4_P(A) = 1/3 - 2/3 = -1/3$, and so $S^4_P(A) < 0$. □

Proposition 23. [POS5] *If $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0$, then $S(A) > 0$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$, and P be a probability distribution where $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.4$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi \mid \psi) = 1/5$, $P(\psi \mid \phi) = 1/5$, $P(\phi \mid \neg\psi) = 4/5$, and $P(\psi \mid \neg\phi) = 4/5$. Therefore $P(\mathsf{C}(A)|\mathsf{E}(A)) > 0$ but $S^i_P(A) < 0$ for $i \in \{1, 2, 3, 4\}$. □

Proposition 24. [POS6] *If $P(\mathsf{E}(A)|\mathsf{C}(A)) > 0$, then $S(A) > 0$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$, and P be a probability distribution where $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.4$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi \mid \psi) = 1/5$, $P(\psi \mid \phi) = 1/5$, $P(\phi \mid \neg\psi) = 4/5$, and $P(\psi \mid \neg\phi) = 4/5$. Therefore $P(\mathsf{E}(A)|\mathsf{C}(A)) > 0$ but $S^i_P(A) < 0$ for $i \in \{1, 2, 3, 4\}$. □

Proposition 25. [NEG1] *If $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$, then $S(A) \leq 0$. Holds for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Assume $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ (S^1_P) So $P(\mathsf{C}(A) \mid \mathsf{E}(A))$ is zero or undefined. So $P(\mathsf{C}(A) \mid \neg\mathsf{E}(A)) \geq 0$ or is undefined. So $S^1_P(A) \leq 0$. (S^2_P) So $P(\mathsf{E}(A) \mid \mathsf{C}(A))$ is zero or undefined, and $P(\mathsf{E}(A) \mid \neg\mathsf{C}(A)) \geq 0$ or is undefined. So $S^2_P(A) \leq 0$. (S^3_P) So $P(\mathsf{C}(A) \mid \mathsf{E}(A))$ is zero or undefined, and $P(\neg\mathsf{C}(A) \mid \mathsf{E}(A))$ is 1 or undefined. So $S^3_P(A) \leq 0$. (S^4_P) So $P(\mathsf{E}(A) \mid \mathsf{C}(A))$ is zero or undefined, and $P(\neg\mathsf{E}(A) \mid \mathsf{C}(A))$ is 1 or undefined. So $S^4_P(A) \leq 0$. □

Proposition 26. [NEG2] *If $P(\mathsf{C}(A)|\mathsf{E}(A)) = 0$, then $S(A) \leq 0$. Holds for S^1_P, S^2_P, S^3_P, and S^2_P.*

Proof. Assume $P(\mathsf{C}(A) \mid \mathsf{E}(A)) = 0$ (S^1_P) So $P(\mathsf{C}(A) \mid \neg\mathsf{E}(A)) \geq 0$ or is undefined. So $S^1_P(A) \leq 0$. (S^2_P) So $P(\mathsf{C}(A) \wedge \mathsf{E}(A)) = 0$ and $P(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) > 0$. So $P(\mathsf{E}(A) \mid \mathsf{C}(A)) = 0$ or is undefined, and $P(\mathsf{E}(A) \mid \neg\mathsf{C}(A)) \geq 0$. So $S^2_P(A) \leq 0$. (S^3_P) So

$P(\neg \mathsf{C}(A) \mid \mathsf{E}(A)) = 1$. So $S^3_P(A) = -1$. (S^4_P) So $P(\mathsf{E}(A) \mid \mathsf{C}(A)) = 0$ or is undefined, and $P(\neg \mathsf{E}(A) \mid \mathsf{C}(A)) = 1$ or is undefined. So $S^4_P(A) \leq 0$. □

Proposition 27. [NEG3] *If $S(A) < 0$, then $P(\mathsf{C}(A)|\mathsf{E}(A)) = 0$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. Counterexamples where $S(A) < 0$ and $P(\mathsf{C}(A)|\mathsf{E}(A))) > 0$. ($S^1_P$ and S^2_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.8$, and $P(\neg\phi \wedge \psi) = 0.1$. So $P(\phi \mid \psi) = 1/2$, $P(\phi \mid \neg\psi) = 1$, $P(\psi \mid \phi) = 1/9$, and $P(\psi \mid \neg\phi) = 1$. So $S^1_P(A) = -1/2$ and $S^2_P(A) = -8/9$. (S^3_P) Let $P(\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \psi) = 0.9$. So $P(\phi \mid \psi) = 1/10$, and $P(\neg\phi \mid \psi) = 9/10$. So $S^3_P(A) = -8/10$. (S^4_P) Let $P(\phi \wedge \psi) = 0.1$, and $P(\phi \wedge \neg\psi) = 0.9$. So $P(\phi \mid \psi) = 1$, $P(\psi \mid \phi) = 1/10$, and $P(\neg\psi \mid \phi) = 9/10$. So $S^4_P(A) = -8/10$. □

Proposition 28. [NEG4] *If $S(A) < 0$, then $P(\mathsf{E}(A)|\mathsf{C}(A))) = 0$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. Let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. Counterexamples where $S(A) < 0$ and $P(\mathsf{E}(A)|\mathsf{C}(A))) > 0$. ($S^1_P$ and S^2_P) Let $P(\phi \wedge \psi) = 0.1$, $P(\phi \wedge \neg\psi) = 0.8$, and $P(\neg\phi \wedge \psi) = 0.1$. So $P(\phi \mid \psi) = 1/2$, $P(\phi \mid \neg\psi) = 1$, $P(\psi \mid \phi) = 1/9$, and $P(\psi \mid \neg\phi) = 1$. So $S^1_P(A) = -1/2$ and $S^2_P(A) = -8/9$. (S^3_P) Let $P(\phi \wedge \psi) = 0.1$, and $P(\neg\phi \wedge \psi) = 0.9$. So $P(\phi \mid \psi) = 1/10$, $P(\neg\phi \mid \psi) = 9/10$, and $P(\psi \mid \phi) = 1$. So $S^3_P(A) = -8/10$. (S^4_P) Let $P(\phi \wedge \psi) = 0.1$, and $P(\phi \wedge \neg\psi) = 0.9$. So $P(\psi \mid \phi) = 1/10$, and $P(\neg\psi \mid \phi) = 9/10$. So $S^4_P(A) = -8/10$. □

Proposition 29. [NEG5] *If $P(\mathsf{C}(A)|\mathsf{E}(A)) < P(\mathsf{C}(A)$, then $S(A) < 0$. Holds for S^1_P and S^2_P, and fails for S^3_P, and S^4_P.*

Proof. Assume $P(\mathsf{C}(A)|\mathsf{E}(A)) < P(\mathsf{C}(A)$. For counterexamples, let A be an argument, where $\mathsf{C}(A) = \phi$ and $\mathsf{E}(A) = \psi$. (S^1_P) By Lemma 1, $P(\mathsf{C}(A)|\mathsf{E}(A)) < P(\mathsf{C}(A)|\neg\mathsf{E}(A))$. So $P(\mathsf{C}(A)|\mathsf{E}(A)) - P(\mathsf{C}(A)|\neg\mathsf{E}(A)) < 0$. So $S^1_P(A) < 0$. (S^2_P) By Lemma 1, $P(\mathsf{C}(A)|\mathsf{E}(A)) < P(\mathsf{C}(A)|\neg\mathsf{E}(A))$. Then by Lemma 3, $P(\mathsf{E}(A)|\mathsf{C}(A)) < P(\mathsf{E}(A)|\neg\mathsf{C}(A))$. So $P(\mathsf{E}(A)|\mathsf{C}(A)) - P(\mathsf{E}(A)|\neg\mathsf{C}(A)) < 0$. So $S^2_P(A) < 0$. (S^3_P) Let $P(\phi \wedge \psi) = 0.3$, $P(\phi \wedge \neg\psi) = 0.4$, $P(\neg\phi \wedge \psi) = 0.2$, and $P(\neg\phi \wedge \neg\psi) = 0.1$. So $P(\phi|\psi) = 3/5$, $P(\phi) = 7/10$, and $S^3_P(A) = 1/5$. (S^4_P) Let $P(\phi \wedge \psi) = 0.4$, $P(\phi \wedge \neg\psi) = 0.3$, $P(\neg\phi \wedge \psi) = 0.3$, and $P(\neg\phi \wedge \neg\psi) = 0$. So $P(\phi|\psi) = 4/7$, $P(\phi) = 7/10$, $P(\psi|\phi) = 4/7$, and $P(\neg\psi|\phi) = 3/7$. So $S^4_4(A) = 1/7$. □

Proposition 30. [INC1] *If $\{\mathsf{C}(A), \mathsf{C}(B)\} \vdash \bot$, then $S(A) < 0$ or $S(B) < 0$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. For a counterexample, consider arguments A where $\mathsf{E}(A) = \psi$ and $\mathsf{C}(A) = \phi$ and B where $\mathsf{E}(B) = \psi'$ and $\mathsf{C}(B) = \neg\phi$. Let P be such that $P(\phi \wedge \psi \wedge \neg\psi') = 0.5$ and $P(\neg\phi \wedge \neg\psi \wedge \psi') = 0.5$. So $P(\phi \mid \psi) = 1$, $P(\phi \mid \neg\psi) = 0$, $P(\psi \mid \phi) = 1$, $P(\psi \mid \neg\phi) = 0$, $P(\neg\phi \mid \psi') = 1$, $P(\neg\phi \mid \neg\psi') = 0$, $P(\psi' \mid \neg\phi) = 1$, and $P(\psi' \mid \phi) = 0$. So $S^i_P(A) > 0$ and $S^i_P(B) > 0$ for $i \in \{1, 2, 3, 4\}$. □

Proposition 31. [INC2] *If $\{\mathsf{C}(A), \mathsf{E}(B)\} \vdash \bot$, then $S(A) < 0$ or $S(B) < 0$. Fails for S^1_P, S^2_P, S^3_P, and S^4_P.*

Proof. For a counterexample, consider arguments A where $\mathsf{E}(A) = \psi$ and $\mathsf{C}(A) = \phi$ and B where $\mathsf{E}(B) = \psi'$ and $\mathsf{C}(B) = \phi'$. Let P be such that $P(\phi \wedge \psi \wedge \neg\phi' \wedge \neg\psi') = 0.5$ and $P(\neg\phi \wedge \neg\psi \wedge \phi' \wedge \psi') = 0.5$. So $P(\phi \mid \psi) = 1$, $P(\phi \mid \neg\psi) = 0$, $P(\psi \mid \phi) = 1$, $P(\psi \mid \neg\phi) = 0$, $P(\phi' \mid \psi') = 1$, $P(\neg\phi' \mid \neg\psi') = 0$, $P(\psi' \mid \phi') = 1$, and $P(\psi' \mid \neg\phi') = 0$. So $S^i_P(A) > 0$ and $S^i_P(B) > 0$ for $i \in \{1, 2, 3, 4\}$. □

Proposition 32. *If argument A is compositionally strong w.r.t. strength measure S^x_P, where $x \in \{1, 2, 3, 4\}$, and threshold $\tau \in [1, -1]$, then the strength of any rule $\rho \in \mathsf{Support}(A)$ is greater than or equal to τ.*

Proof. An argument A is compositionally strong with respect to strength measure S^x_P and threshold $\tau \in [-1, 1]$ iff for all $B \in \mathsf{Intermediates}(A)$, $S^x_P(B) \geq \tau$. So for all $B \in \mathsf{Intermediates}(A)$ of the form $B = \langle\{\psi_1, \ldots, \psi_n, \psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi\}, \phi\rangle$, $S^x_P(B) \geq \tau$. So for all rules $\rho = \psi_1 \wedge \ldots \wedge \psi_n \rightarrow \phi \in \mathsf{Support}(A)$, the strength of ρ is greater than or equal to τ. □

Proposition 33. *For an argument A and a probability distribution P, if P^* is the probability distribution obtained by Bayes' rule (i.e. $P^*(\mathsf{C}(A)) = P(\mathsf{C}(A)|\mathsf{E}(A))$), then*

- $S^1_{P^*}(A) = P^*(\mathsf{C}(A))$
- $S^2_{P^*}(A) = 0$
- $S^3_{P^*}(A) = P^*(\mathsf{C}(A)) - P^*(\neg\mathsf{C}(A))$
- $S^4_{P^*}(A) = 1$

Proof. We assume $P^*(\mathsf{E}(A)) = 1$, and so $P^*(\mathsf{C}(A) \wedge \mathsf{E}(A)) = P^*(\mathsf{C}(A))$, $P^*(\neg\mathsf{C}(A) \wedge \mathsf{E}(A)) = P^*(\neg\mathsf{C}(A))$, $P^*(\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$, and $P^*(\neg\mathsf{C}(A) \wedge \neg\mathsf{E}(A)) = 0$. Therefore, $P^*(\mathsf{C}(A)|\mathsf{E}(A)) = P^*(\mathsf{C}(A) \wedge \mathsf{E}(A))/P^*(\mathsf{E}(A)) = P^*(\mathsf{C}(A))$, $P^*(\mathsf{C}(A)|\neg\mathsf{E}(A))$ is undefined, $P^*(\mathsf{E}(A)|\mathsf{C}(A)) = P^*(\mathsf{C}(A))/P^*(\mathsf{C}(A)) = 1$ and $P^*(\mathsf{E}(A)|\neg\mathsf{C}(A)) = P^*(\neg\mathsf{C}(A))/P^*(\neg\mathsf{C}(A)) = 1$. Therefore we get the following: (1) $S^1_{P^*}(A) = P^*(\mathsf{C}(A))$; (2) $S^2_{P^*}(A) = 0$; (3) $S^3_{P^*}(A) = P^*(\mathsf{C}(A)) - P^*(\neg\mathsf{C}(A))$; and (4) $S^4_{P^*}(A) = 1$. □

References

[1] L. Amgoud, J. Ben-Naim, Axiomatic foundations of acceptability semantics, in: Proceedings of KR'16, AAAI Press, 2016, pp. 2–11.
[2] T. Bench-Capon, S. Doutre, P. Dunne, Audiences in argumentation frameworks, Artif. Intell. 171 (1) (2007) 42–71.
[3] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A comparative study of ranking-based semantics for abstract argumentation, in: Proceedings of AAAI'16, 2016.
[4] T. Bench-Capon, Audiences and argument strength, in: Proceedings of the 3rd Workshop on Argument Strength, FernUniversität in Hagen, Germany, 2021.
[5] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, Artif. Intell. 171 (2007) 675–700.
[6] Ph. Besnard, A. García, A. Hunter, S. Modgil, H. Prakken, G. Simari, F. Toni, Introduction to structured argumentation, Argum. Comput. 5 (1) (2014) 1–4.
[7] Ph. Besnard, A. Hunter, A review of argumentation based on deductive arguments, in: Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, Leendert van der Torre (Eds.), Handbook of Formal Argumentation, vol. 1, College Publications, 2018, pp. 435–482.
[8] M. Beirlaen, J. Heyninck, P. Pardo, C. Straßer, Argument strength in formal argumentation, J. Appl. Log. 5 (3) (2018) 629–676.
[9] G. Boole, An Investigation of the Laws of Thought: On Which Are Founded the Mathematical Theories of Logic and Probabilities, Walton and Maberly, 1854.
[10] P. Baroni, A. Rago, F. Toni, From fine-grained properties to broad principles for gradual argumentation: a principled spectrum, Int. J. Approx. Reason. 105 (2019) 252–286.
[11] R. Carnap, Logical Foundations of Probability, 2nd edition, University of Chicago Press, 1962.
[12] A. Cohen, S. Gottifredi, L. Tamargo, A. García, G. Simari, An informant-based approach to argument strength in defeasible logic programming, Argum. Comput. 12 (1) (2021) 115–147.
[13] P. Cheeseman, A method of computing generalized bayesian probability values for expert systems, in: Proceedings of IJCAI'83, AAAI Press, 1983, pp. 198–202.
[14] D. Christensen, Measuring confirmation, J. Philos. 96 (1999) 437–461.
[15] C. Cayrol, M.-C. Lagasquie-Schiex, Graduality in argumentation, J. Artif. Intell. Res. 23 (2005) 245–297.
[16] K. Cyras, A. Rago, E. Albini, P. Baroni, F. Toni, Argumentative XAI: a survey, in: Proceedings of IJCAI'21, ijcai.org, 2021, pp. 4392–4399.
[17] V. Crupi Confirmation, in: The Stanford Encyclopedia of Philosophy, Spring 2020 edition, Metaphysics Research Lab, Stanford University, 2020.
[18] P. Dellunde, L. Godo, A. Vidal, Probabilistic argumentation: an approach based on conditional probability -a preliminary report, in: Proceedings of JELIA'21, in: Lecture Notes in Computer Science, vol. 12678, Springer, 2021, pp. 25–32.
[19] P.M. Dung, P.M. Thang, Towards (probabilistic) argumentation for jury-based dispute resolution, in: Proceedings of COMMA'10, IOS Press, 2010, pp. 171–182.
[20] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (2) (1995) 321–358.
[21] D. Etherington, R. Reiter, On inheritance hierarchies with exceptions, in: Proc. of AAAI'83, AAAI Press, 1983, pp. 104–108.
[22] G. Governatori, M. Maher, G. Antoniou, D. Billington, Argumentation semantics for defeasible logic, J. Log. Comput. 14 (5) (2004) 675–702.
[23] M. Goldszmidt, P. Morris, J. Pearl, A maximum entropy approach to nonmonotonic reasoning, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1993) 220–232.
[24] A. Garcia, Guillermo R. Simari, Defeasible logic programming: an argumentative approach, Theory Pract. Log. Program. 4 (1–2) (2004) 95–138.
[25] R. Haenni, Modelling uncertainty with propositional assumptions-based systems, in: Applications of Uncertainty Formalisms, in: LNCS, vol. 1455, Springer, 1998, pp. 446–470.
[26] A. Hunter, L. Chalaguine, T. Czernuszenko, E. Hadoux, S. Polberg, Towards computational persuasion via natural language argumentation dialogues, in: Proceedings of KI'19, in: LNCS, vol. 11793, Springer, 2019, pp. 18–33.
[27] E. Hadoux, A. Hunter, Computationally viable handling of beliefs in arguments for persuasion, in: Proceedings of ICTAI'16, IEEE Computer Society, 2016, pp. 319–326.
[28] E. Hadoux, A. Hunter, Comfort or safety? gathering and using the concerns of a participant for better persuasion, Argum. Comput. 10 (2) (2019) 113–147.
[29] E. Hadoux, A. Hunter, J.-B. Corrégé, Strategic dialogical argumentation using multi-criteria decision making with application to epistemic and emotional aspects of arguments, in: Proceedings of FoIKS'18, in: Lecture Notes in Computer Science, vol. 10833, Springer, 2018, pp. 207–224.
[30] A. Hunter, S. Polberg, Empirical methods for modelling persuadees in dialogical argumentation, in: Proceedings of ICTAI'17, IEEE Computer Society, 2017, pp. 382–389.
[31] A. Hunter, N. Potyka, Updating probabilistic epistemic states in persuasion dialogues, in: Proceedings of ECSQARU'17, in: LNCS, vol. 10369, Springer, 2017, pp. 46–56.
[32] A. Hunter, M. Thimm, Probabilistic reasoning with abstract argumentation frameworks, J. Artif. Intell. Res. 59 (2017) 565–611.
[33] A. Hunter, A probabilistic approach to modelling uncertain logical arguments, Int. J. Approx. Reason. 54 (1) (2013) 47–81.
[34] A. Hunter, Generating instantiated argument graphs from probabilistic information, in: Proceedings of ECAI'20, IOS Press, 2020.
[35] A. Hunter, Argument strength in probabilistic argumentation using confirmation theory, in: Proceedings of ECSQARU'21, in: LNCS, vol. 2000, Springer, 2021, pp. 100–110.
[36] R. Jeffrey, The Logic of Decision, Chicago University Press, 1965.
[37] J. Kemeny, P. Oppenheim, Degrees of factual support, Philos. Sci. 19 (1952) 307–324.
[38] S. Lukin, P. Anand, M. Walker, S. Whittaker, Argument Strength Is in the Eye of the Beholder: Audience Effects in Persuasion, Proceedings of EACL'17, vol. 1, Association for Computational Linguistics, 2017, pp. 742–753.
[39] H. Li, N. Oren, T.J. Norman, Probabilistic argumentation frameworks, in: Proceedings of TAFA'11, in: LNCS, vol. 7132, Springer, 2011, pp. 1–16.
[40] J. McCarthy, Circumscription: a form of non-monotonic reasoning, Artif. Intell. 28 (1) (1980) 89–116.
[41] S. Modgil, H. Prakken, The ASPIC+ framework for structured argumentation: a tutorial, Argum. Comput. 5 (2014) 31–62.
[42] P.-A. Matt, F. Toni, A game-theoretic perspective on the notion of argument strength in abstract argumentation, in: Proceedings of JELIA'08, in: LNCS, Springer, 2008, pp. 285–297.
[43] N. Nilsson, Probabilistic logic, Artif. Intell. 28 (1995) 71–87.
[44] R. Nozick, Philosophical Explanations, Oxford University Press, 1981.
[45] J.B. Paris, The Uncertain Reasoner's Companion – A Mathematical Perspective, Cambridge University Press, 1994.
[46] N. Pfeifer, On argument strength, in: Bayesian Argumentation, in: Synthese Library, vol. 362, Springer, 2012, pp. 185–193.

[47] S. Polberg, A. Hunter, Empirical evaluation of abstract argumentation: supporting the need for bipolar and probabilistic approaches, Int. J. Approx. Reason. 93 (2018) 487–543.
[48] S. Polberg, A. Hunter, M. Thimm, Belief in attacks in epistemic probabilistic argumentation, in: Proceedings of SUM'17, in: LNCS, vol. 10564, Springer, 2017, pp. 223–236.
[49] J. Pollock, Defeasible reasoning, Cogn. Sci. 11 (4) (1987) 481–518.
[50] J. Pollock, Cognitive Carpentry, MIT Press, 1995.
[51] N. Pfeifer, H. Pankka, Modeling the ellsberg paradox by argument strength, in: Proceedings of the Cognitive Science Society (CogSci'17), cognitivesciencesociety.org, 2017.
[52] H. Prakken, Probabilistic strength of arguments with structure, in: Proceedings of KR'18, AAAI Press, 2018, pp. 158–167.
[53] H. Prakken, Philosophical reflections on argument strength and gradual acceptability, in: Proceedings of ECSQARU'21, in: Lecture Notes in Computer Science, vol. 12897, Springer, 2021, pp. 144–158.
[54] J. Raab, Troubles with Bayesian argumentation, in: Proceedings of the 3rd Workshop on Argument Strength, FernUniversität in Hagen, Germany, 2021.
[55] R. Riveret, G. Governatori, On learning attacks in probabilistic abstract argumentation, in: Proceedings of AAMAS'16, 2016, pp. 653–661.
[56] R. Riveret, A. Rotolo, G. Sartor, H. Prakken, B. Roth, Success chances in argument games: a probabilistic approach to legal disputes, in: Proceedings of JURIX'07, IOS Press, 2007, pp. 99–108.
[57] G. Simari, R. Loui, A mathematical treatment of defeasible reasoning and its implementation, Artif. Intell. 53 (2–3) (1992) 125–157.
[58] P. Shakarian, G. Simari, G. Moores, D. Paulo, S. Parsons, M. Falappa, A. Aleali, Belief revision in structured probabilistic argumentation - model and application to cyber security, Ann. Math. Artif. Intell. 78 (3–4) (2016) 259–301.
[59] K. Tentori, V. Crupi, N. Bonini, D. Osherson, Comparison of confirmation measures, Cognition 103 (2007) 107–119.
[60] M. Thimm, A probabilistic semantics for abstract argumentation, in: Proceedings of ECAI'12, 2012.
[61] F. Toni, A tutorial on assumption-based argumentation, Argum. Comput. 5 (1) (2014) 89–117.
[62] B. Verheij, Arguments and their strength: revisiting Pollock's anti-probabilistic starting points, in: Proceedings of COMMA'14, IOS Press, 2014.