Argument & Computation 14 (2023) 75–107
DOI 10.3233/AAC-210016
IOS Press

Probabilistic interpretations of argumentative attacks: Logical and experimental results[1]

Niki Pfeifer [a,*] and Christian G. Fermüller [b]
[a] *Department of Philosophy, University of Regensburg, Universitätsstraße 31, 93040 Regensburg, Germany*
E-mail: niki.pfeifer@ur.de
[b] *Institute of Logic and Computation, TU Wien, Favoritenstraße 9–11, 1040 Vienna, Austria*
E-mail: chrisf@logic.at

Abstract. We present an interdisciplinary approach to argumentation combining logical, probabilistic, and psychological perspectives. We investigate logical attack principles which relate attacks among claims with logical form. For example, we consider the principle that an argument that attacks another argument claiming A triggers the existence of an attack on an argument featuring the stronger claim $A \wedge B$. We formulate a number of such principles pertaining to conjunctive, disjunctive, negated, and implicational claims. Some of these attack principles seem to be *prima facie* more plausible than others. To support this intuition, we suggest an interpretation of these principles in terms of coherent conditional probabilities. This interpretation is naturally generalized from qualitative to quantitative principles. Specifically, we use our probabilistic semantics to evaluate the rationality of principles which govern the strength of argumentative attacks. In order to complement our theoretical analysis with an empirical perspective, we present an experiment with students of the TU Vienna ($n = 139$) which explores the psychological plausibility of selected attack principles. We also discuss how our qualitative attack principles relate to well-known types of logical argumentation frameworks. Finally, we briefly discuss how our approach relates to the computational argumentation literature.

Keywords: Argumentative attacks, coherence, psychological experiment, logical argumentation, probabilistic reasoning, rationality principles

1. Introduction

Various disciplines study argumentation [8,85], including artificial intelligence (e.g., [10,75]), computer science (e.g., [11,30]), philosophy (e.g., [34,52,84,86]), and psychology (e.g., [50,62]). The motivation of this paper is to bring together logical, probabilistic, and psychological points of views to better understand specific rationality principles, which refer to the logical form of claims, but ignore the support part of arguments. Our approach is hence an interdisciplinary one, as we combine elements of Dung-style abstract argumentation [30], logical argument forms, coherent conditional probability, and also present a psychological experiment to assess the descriptive validity of selected formal principles.

Argumentation is a highly complex and dynamic process that proceeds dialectically by presenting arguments and counter-arguments, i.e., attacks on arguments. Like in Dung-style abstract argumentation [30], we take a static view that ignores temporal aspects of argumentation and focuses on the attack

[1] This work is a substantially expanded version of a contribution presented at the *11th Workshop on Uncertainty Processing* (WUPES 2018), Třeboň, Czech Republic.

[*] Corresponding author. E-mail: niki.pfeifer@ur.de.

relation. Usually, arguments are conceived as premise ("support") and conclusion ("claim") pairs. Here, following [21–23], we focus on the interplay between argumentative attacks and the logical form of claims formalized by classical propositional formulæ. Since we ignore the support part of arguments, our attack relation operates not between arguments, but between propositions (claims of arguments). This relation can be understood as the result of an existential abstraction: a claim A attacks a claim B, if there exists an argument with claim A that attacks an argument with claim B in an underlying instantiated argumentation framework. The term 'semi-abstract argumentation framework (SAF)' was coined in [22] to emphasize the fact that corresponding attack principles operate on a level that is situated between Dung's (fully) abstract argumentation frameworks and (fully) instantiated argumentation framework. This corresponds to the claim-centered view on argumentation [35]; SAFs are called 'claim augmented argumentation frameworks' in [35].

In [22] logical attack principles have been introduced that are motivated by considerations like the following: if an argumentation framework contains arguments that feature claims A, B, as well as $A \wedge B$, respectively, then it seems reasonable to expect that for any argument that attacks an argument with claim A or B there is also an argument attacking an argument with claim $A \wedge B$. However, it is much less clear, whether one is also entitled to expect an attack against an argument with either claim A or with claim B if there exists an argument attacking an argument with claim $A \wedge B$. In [23] such (qualitative) logical attack principles where generalized to quantitative principles, where the attack relation between claims is endowed with weights in [0, 1]. For example, the following principle was considered there: if there is an attack with weight x on an argument claiming A and an attack with weight y on an argument claiming A, then an attack against an argument with claim $A \wedge B$ should carry a weight $\geqslant \max(x, y)$. Both, the qualitative and the quantitative scenario, call for a systematic assessment of logical attack principles of the indicated type. The distinguishing feature of our paper is that we endow qualitative as well as quantitative versions of logical attack principles with a probabilistic interpretation that allows us to distinguish between plausible and implausible forms in a principled way, which we also assess empirically. Let us also clarify at the outset that we are interested in principles here that are independent of the concrete content of arguments, but rather only refer to the logical form of the involved claims. A similar proviso applies to weights of attack: we neither propose any particular method of assigning weights nor impose any particular meaning of attack strength that may depend on the given context. Rather we suggest and evaluate principles that potentially apply to any (normalizable) notion of strength of attack.

The outline of the paper is as follows: Section 2 explains the specific level of abstraction of our approach. Section 3 gives a brief survey of qualitative attack principles which were investigated in [22]. We propose a probabilistic interpretation of attack between claims. Specifically, we use coherent conditional probabilities to systematically evaluate the rationality of logical attack principles: coherence serves as rationality criterion for selecting attack principles: "good" principles should be coherent, i.e., they should not violate laws of probability. In Section 4 we show how to model the qualitative attack principles in probabilistic terms. An attractive feature of our probabilistic semantics is that it naturally leads to an interpretation of weighted attacks. The corresponding generalization of qualitative to quantitative attack principles and their probabilistic interpretation is discussed in Section 5. Section 6 presents an experiment which aims to explore the psychological plausibility of selected features of the proposed approach. Then, we present in Section 7 some observations concerning the special case of logical argumentation, where the underlying attack relation is defined in terms of classical logical entailment. Section 8 contextualizes our contributions by indicating some relations to other approaches in computational argumentation and AI. We conclude in Section 9 with some remarks on future research.

2. Argumentation frameworks: Abstract, concrete, and semi-abstract

Dung's seminal paper [30] introduced abstract argumentation frameworks (AFs). An AF is a directed graph, where the vertices represent arguments and the edges represent attacks between these arguments. Given an AF, the primary task is to compute extensions or admissible sets; i.e., sets of arguments that don't attack each other and that moreover defend the arguments in the extension by attacking those arguments outside the extension that attack them. Various further properties, in particular maximality conditions, imposed on extensions lead to a plethora of so-called semantics for AFs, including complete, preferred, grounded, and stable semantics. We will not be concerned with these types of extensions here and refer the interested reader to, e.g., the handbook [8] for details.

Dung and his followers showed that the indicated lean and mathematically elegant abstract approach, based on graph theoretic properties of the attack relation, allows one to computationally handle various reasoning tasks arising for nonmonotonic reasoning, including, e.g., forms of logic programming. It is indeed impressive to observe to what extent purely structural properties of graphs (AFs), compiled from large, in general inconsistent sets of statements, assist the assessment of information entailed by such data bases. However, it has also been recognized that one has to pay attention to the logical structure of arguments themselves in order to be able to determine whether a given argument indeed attacks another argument or not. Various logical formats for arguments have been suggested in the literature. For example, Besnard and Hunter [11] have popularized a widely followed approach in which arguments are conceived as pairs $\langle \Phi, A \rangle$, where the *support* Φ consists of a finite, consistent set of formulæ entailing the *conclusion* or *claim* A. Moreover, Φ is required to be *minimal* (with respect to the subset relation) among sets of formulas with these properties. Other authors (e.g., [7,25,48]) have argued that both, the consistency as well as the minimality condition, are problematic. In particular, Arieli and Strasser [6,83] explore sequent-based argumentation, where arguments are identified with (single-conclusioned) sequents $\Phi \vdash A$. In this approach the support part Φ, i.e. the formulas on the left hand side of the sequent, neither needs to be consistent nor minimal. Yet another quite popular approach is ASPIC$^+$ [61], where the structure of arguments is more involved, featuring not only formulas expressing facts, but also default rules as well as strict (logical) rules in the support part of arguments. All of the mentioned formats refer to *logical argumentation*, where the claim of an argument has to be logically entailed by its support. Moreover, also the attack relation between arguments is defined in terms of logical consequence in various ways. While this is in line with the indicated computational approach to argumentation, logical argumentation is arguably too restrictive to support realistic models of informal argumentation, where attack between arguments is, in general, not a logical relation, but a material one that depends on given interpretations and contexts and that might admit degrees. Although the attack principles that are in the focus of this paper refer to the logical form of claims of arguments, they are not confined to logical argumentation. In particular, except for the specific remarks on logical argumentation in Section 7, we will not be concerned with the specific type of attack that relates two arguments.[2] For the logical attack principles introduced in [22] and described in Section 3, below, the specific form of attack is immaterial. In fact, these principles amount to (possible) rationality constraints for AFs also for collections of arguments, where the attack relation between pairs of arguments is not of a logical nature at all.

As outlined above, Dung-style argumentation theory can be thought of as referring to two quite different levels. On the one hand, there are the abstract AFs, where arguments are represented simply as nodes

[2] Likewise, we are not concerned with the nature of the relation between the claim and the support of argument. But note that any argument, where the claim A is not logically entailed by its support $\{B_1, \ldots, B_n\}$, may be turned into a logical argument by adding the formula $(B_1 \wedge \cdots \wedge B_n) \supset A$ to its support.

in a directed graph and edges between nodes represent attacks between arguments. On the other hand, there are concrete (instantiated) AFs, where arguments are structured compounds of specific logically complex statements and, possibly, rules of different kinds. The logical attack principles, introduced in [22], that we study in this paper neither operate on abstract AFs nor on the level of concrete AFs. These principles rather focus on the *logical form* of claims, i.e. on the outermost logical connective of the formula representing the claim of an argument. A particularly simple example of an attack principle of this kind is the following: if an argument γ attacks an argument α that features a claim A, then γ implicitly also attacks an argument β, if the claim of β is $A \wedge B$. (Actually, as we will see in Section 3, the attack principles considered in this paper are somewhat less restrictive: rather than requiring that γ itself attacks β, the principle is satisfied if there exists an argument that attacks β.)

Formally, following [22,23], we thus consider *semi-abstract AFs*, which are ordinary AFs, where each node is annotated with a propositional formula featuring the claim of the represented argument. The expression 'semi-abstract' is meant to signal that we are not interested in the possibly quite complex internal structure of concrete arguments; rather, we add information about the logical form of claims to the abstract AF. The emphasis on claims of arguments, rather than on full arguments is by no means new, of course. In fact, the standard approach to computational argumentation in the wake of Dung features the focus on claims as its final step of information extraction: once the appropriate (e.g., complete, preferred, grounded, or stable) extensions have been computed for a given collection of arguments, one usually asks whether a given formula appears as claim A of some argument in either all or at least in some extensions. In the former case, A is *skeptically accepted*; in the latter case A is *credulously accepted*. In the context of investigating the computational complexity of corresponding reasoning tasks, Dvořák and Woltran [35] speak of a *claim-centric view* on AFs. Although we aim at a different target, namely the interpretation of certain rationality constraints on the attack relation, our investigation is claim-centric as well. Note that what has been called semi-abstract AF (SAF) in [22] is called claim-augmented AF (CAF) in [35].

It may seem obvious that in order to attack (an argument with claim) $A \wedge B$ it suffices to attack either A or B. But it is much less clear whether also the inverse holds, i.e., whether attacking $A \wedge B$ entails attacking A or attacking B. Neither is it immediately obvious which principles of this type are plausible for disjunctive, implicational, and negated claims. The situation gets even more challenging when, as in [23], one generalizes from semi-abstract AFs to *weighted* semi-abstract AFs, where a weight is attached to each edge of the graph, intended to signal the strength of the corresponding attack. Still, it seems intuitively justified to impose constraints like the following: if an argument γ attacks an argument with claim $A \vee B$ with a certain degree of strength, then either γ itself or some related argument in the given framework attacks arguments with claims A and B, respectively, to at least the same degree. In order to assess the plausibility of this and similar principles in a systematic manner one needs a concrete *interpretation* of the notion of (weighted) attacks between claims of arguments. After introducing a range of candidates for qualitative (i.e., unweighted) attack principles in the next section, we will provide a probability-based interpretation for them in Section 4. This interpretation straightforwardly generalizes to quantitative principles (i.e., involving weighted attacks) as will be shown in Section 5.

3. Qualitative attack principles

Following [22], we write "$A \longrightarrow B$" to denote that there is an argument claiming A that attacks some argument with the claim B. This notion implicitly refers to a given collection Λ of arguments (i.e., a

given AF). However, we neither care about the particular form of the arguments in Λ nor about the nature of the attack relation (defeat, rebuttal, undercut, etc.) defined for Λ. Rather, we abstract away from the given arguments and corresponding attacks and focus on formulas that appear as claims of arguments. Therefore we can safely drop the reference to Λ. Although, strictly speaking, arguments and not claims get attacked, we will express $A \longrightarrow B$ as "A attacks B", which, as just explained, is to be understood as short for "there exists an argument α with claim A in Λ that attacks at least one argument β in Λ with claim B".

From a computational point of view, one may think of 'semi-abstraction' as follows: given Λ one compiles a graph, called semi-abstract argumentation framework (SAF) in [22,23], where the set of vertices is the set of formulas that appear as claims in arguments in Λ. The edges of the SAF are readily computed as indicated above: whenever an argument with claim A attacks an argument with claim B, then there is an edge $A \longrightarrow B$. Clearly, extracting an SAF from an underlying instantiated argumentation framework Λ can be done in polynomial time. The attack principles investigated below refer to the extracted SAF, rather than the underlying AF. In general, the AF is much larger than the corresponding SAF. In any case, checking whether an argumentation framework satisfies logical attack principles, like those presented below, is decidable in polynomial time. This should be contrasted with the complexity of checking, e.g., whether a given set of claims is logically consistent (i.e., satisfiable) or with the intractability results regarding preferred or stable semantics (see, e.g., [33]).

Throughout the paper we will refer to the following example where we illustrate some relevant notions by appealing to a meteorological argumentation framework $\mathcal{M}$.

Meteorological Example 1. Suppose that there is a database, containing possibly incomplete and inconsistent meteorological data from which arguments featuring claims referring to the weather of the next day at a particular place get extracted in some manner. The support parts of the arguments in the resulting instantiated AF $\mathcal{M}$ directly refer to the entries in the meteorological database. However, sticking with the paradigm of semi-abstract AFs, we are only interested in the claims of the arguments. For the sake of concreteness let us consider the following two statements.

R**:** "It will rain tomorrow."
S**:** "It will be sunny tomorrow."

In writing $R \longrightarrow S$, we refer to the fact that $\mathcal{M}$ contains at least one argument that claims that it will rain tomorrow that attacks some argument in $\mathcal{M}$ that claims that it will be sunny. (There may be many such arguments. But this is immaterial here.)

Various ways in which concrete meteorological data may support a claim like R (or S) are conceivable. Note, however, that we deliberately abstract away from the particular manner in which the support part of an argument supports its featured claim. Likewise, we will not make any assumptions about the specific manner in which the attack relation between arguments may be defined. For our purpose it is sufficient to assume that it is specified in some formal or informal manner whether a given argument attacks another given argument, or not.[3]

In the following we will assume that the claims of arguments are presented as classical propositional formulas. Using classical logic allows us to identify propositions with events, which we will be used in

[3]We will make an exception to the just outlined approach in Section 7. There we will make a few observations concerning the special case of logical argumentation, where the claim of an argument is assumed to be a logical consequence of its support and where the attack relation can be defined as a logical relation in various ways.

Section 4. It is natural to assume that attacking a claim A triggers an implicit attack to any claim B that classically logically entails A (denoted by $B \models A$):

(C.gen) If $F \longrightarrow A$ and $B \models A$, then $F \longrightarrow B$,

where F, A, and B are arbitrary propositional formulas.[4] Here, **C** indicates *classical logic* and **gen** indicates the *generality* of the principle. However, we will not consider arbitrary pairs of claims, where one is the logical consequence of the other. Rather, we are interested in the relation between logically compound formulas (formed by conjunction $\wedge$, disjunction $\vee$, material conditional $\supset$, and negation $\neg$) and their immediate subformulas. The following principles, called *logical attack principles* in [22, 23], can be seen as instances of the general principle (**C.gen**). (Actually each, (**C.**$\wedge$) as well as (**C.**$\vee$), combine two instances of (**C.gen**).)

(**C.**$\wedge$) If $F \longrightarrow A$ or $F \longrightarrow B$ then $F \longrightarrow A \wedge B$.
(**C.**$\vee$) If $F \longrightarrow A \vee B$ then $F \longrightarrow A$ and $F \longrightarrow B$.
(**C.** $\supset$) If $F \longrightarrow A \supset B$ then $F \longrightarrow B$.

Note that (**C.**$\wedge$) can be replaced by the principle "*If* $F \longrightarrow A$ *then* $F \longrightarrow A \wedge B$", which is equivalent to "*If* $F \longrightarrow B$ *then* $F \longrightarrow A \wedge B$", because of the commutativity of conjunction. For the sake of clarity, we spell out the full meaning of (**C.**$\wedge$) as an example. It refers to some underlying argumentation framework Λ in which one can find arguments with claims A, B, as well as $A \wedge B$, respectively. (Of course, there may be many arguments for each of these claims in Λ.) The principle (**C.**$\wedge$) is satisfied if the following holds: if Λ contains an argument ϕ with claim F, such that ϕ attacks some argument α in Λ that has claim A or such that ϕ attacks some argument β in Λ that has claim B, then Λ contains an argument ϕ' with claim F that attacks an argument γ of Λ featuring the claim $A \wedge B$. (The other principles can be spelled out analogously.)

Concerning negation, the following principle is intuitively plausible.

(**C.**$\neg$) For non-contradictory formulas F: if $F \longrightarrow A$ then $F \not\longrightarrow \neg A$,

where, for arbitrary formulas G and H, $G \not\longrightarrow H$ states that, in the underlying AF, no argument that claims G attacks any argument that claims H. Like for the positive case, we abbreviate this by 'H is not attacked by G' or, equivalently, 'G does not attack H'. The restriction to non-contradictory (i.e., satisfiable) formulas in (**C.**$\neg$) is necessary in light of the guiding general principle (**C.gen**).[5] Since a contradictory formula F entails *every* formula, we expect that every argument with a contradictory claim attacks every argument and hence also $F \longrightarrow \neg A$ for arbitrary claims A. Hence we stipulate F to be non-contradictory.

Meteorological Example 2. Continuing Example 1, let us instantiate the formulas mentioned in the above attack principles with concrete statements as follows.

R**:** "It will rain tomorrow."
S**:** "It will be sunny tomorrow."
W**:** "It will be warm tomorrow."

[4] Since we focus only on the logical form of the attacked claim, we use F as a generic sign throughout the paper for the formula that denotes the claim of the attacking argument.

[5] This restriction is not explicitly stated in [22] and [23] for the principle (**A.**$\neg$) that corresponds to (**C.**$\neg$). With hindsight, this is a problematic omission.

Principle (**C.**$\wedge$) thus gets instantiated to the following possible property of the meteorological AF $\mathcal{M}$. Suppose that $\mathcal{M}$ contains an argument claiming that it will rain tomorrow (R) that attacks an argument claiming that it will be sunny tomorrow (S). Then, under the condition that $\mathcal{M}$ also contains arguments claiming that it will be sunny *and* warm tomorrow ($S \wedge W$), at least one such argument will be attacked by some argument claiming that it will rain tomorrow.

In the same manner one can instantiate also the other logical attack principles. Since (**C.**$\neg$) involves negation on the meta-level of talking about attacks as well as on the object level of claims, it may be helpful to instantiate it explicitly. For the above concrete claims R and S, (**C.**$\neg$) expresses the following (possible) property of $\mathcal{M}$. Suppose that $\mathcal{M}$ contains an argument claiming that it will rain tomorrow that attacks an argument claiming that it will be sunny tomorrow. Then, to satisfy (**C.**$\neg$), $\mathcal{M}$ does not contain arguments that claim that it will rain and that attack the claim that it will not be sunny tomorrow.

One can also formulate inverse forms of the above principles:

(**C.**$\wedge$)**'** If $F \longrightarrow A \wedge B$ then $F \longrightarrow A$ or $F \longrightarrow B$.
(**C.**$\vee$)**'** If $F \longrightarrow A$ and $F \longrightarrow B$ then $F \longrightarrow A \vee B$.
(**C.** $\supset$)**'** If $F \longrightarrow B$ then $F \longrightarrow A \supset B$.
(**C.**$\neg$)**'** For non-contradictory formulas F: if $F \not\longrightarrow A$ then $F \longrightarrow \neg A$.

These last mentioned principles seem, at least partly, to be intuitively much more demanding than those following from (**C.gen**). To get a better feeling for the intuitive (in)plausibility of the principles, let us continue our running example.

Meteorological Example 3. Let us use again refer to our meteorological AF $\mathcal{M}$ and use the same concrete statements for R, S, and W, respectively, as in Example 2. Suppose that $\mathcal{M}$ contains an argument claiming that it will rain tomorrow that attacks an argument claiming that it will be sunny and warm tomorrow ($R \longrightarrow S \wedge W$). Then, to satisfy principle (**C.**$\wedge$)**'**, $\mathcal{M}$ would have to contain an argument claiming that it will rain tomorrow that attacks an argument that either features the claim "It will be sunny tomorrow" or the claim "It will be warm tomorrow". While this concrete instance of (**C.**$\wedge$)**'** is not outright wrongheaded, it amounts to an intuitively much stronger (possible) constraint on the attack relation of $\mathcal{M}$ compared to the corresponding instance of (**C.**$\wedge$) in Example 2. It is at least conceivable that the underlying meteorological data that indicate that it will rain tomorrow are incompatible with a scenario where it will be sunny and warm, without being incompatible with either the forecast that it will be sunny (but cool) or that will be warm (without sunshine). In contrast, principle (**C.**$\wedge$) only reflects a property of conjunction that does not amount to constraints about possible weather scenarios. In other words, differently to (**C.**$\wedge$)**'**, the plausibility of (**C.**$\wedge$) does not depend on the concrete meaning of the involved logically atomic propositions. (Similar consideration hold for the principles (**C.**$\vee$)**'**, (**C.** $\supset$)**'**, and (**C.**$\neg$)**'**, contrasted with (**C.**$\vee$), (**C.** $\supset$), and (**C.**$\neg$).)

The results of [22] imply that imposing *all* of the above (connective specific) attack principles amounts to an alternative characterization of classical logic, while proper subsets of the full set of these principles lead to weaker logics that result from discarding some of the logical inference rules of Gentzen's classical sequent calculus **LK** [40].

Systematic criteria for accepting or rejecting attack principles call for a robust interpretation of the attack relation that is capable of formally supporting (or questioning, as appropriate) informal intuitions

about the varying strength of the attack principles.[6] In the next section we will tackle this problem by applying coherence-based probability theory for developing a semantics of our qualitative attack principles. This will also provide a natural and straightforward basis for investigating quantitative attack principles.

4. Probabilistic semantics

Probabilistic semantics for argumentation became popular in recent years (see, e.g., [49,53,54,63,76,87]). In light of the results of [22], as sketched in Section 3, the challenge is to come up with an intuitively convincing and formally sound interpretation of the attack relation between claims of arguments. This motivates us to explore to which extent one may employ coherence-based *conditional probability* (see, e.g., [20,41,71]) for this purpose. The basic intuition of coherence is usually explained in betting terms, specifically in terms of avoiding Dutch books. Accepting a Dutch book implies sure loss, thus making sure to avoid such bets is the basic rationality requirement.

Definition 1. An assessment on an arbitrary family $\mathcal{C}$ of conditional events is *coherent* if and only if, for any combination of bets on a finite subset of conditional events in $\mathcal{C}$, it cannot happen that the values of the random gain, when at least one bet is not called off, are all positive or all negative.

Coherence amounts to the solvability of a suitable finite sequence of systems of linear equations (for corresponding algorithms to check coherence and for further technical details see, e.g., [12,20]). A *conditional event* $C|A$ is the (conditional, trivalent) object which is measured by the corresponding conditional probability $p(C|A)$.

Definition 2. A *conditional event* $C|A$ is *true* if $A \wedge C$ is true, *false* if $A \wedge \neg C$ is true, and *void* (or *undetermined*) if $\neg A$ is true.

In betting terms, Definition 2 can be read such that you *win* the bet on $C|A$ when $A \wedge C$ is true, you *lose* when $A \wedge \neg C$ is true, and you *get your money back* when $\neg A$ is true. Because of its trivalence, $C|A$ cannot be expressed by any Boolean function. Within the coherence approach, conditional probability is primitive (and not defined by the fraction, $p(A \wedge C)/p(A)$, which—in order to avoid fractions over zero—requires positive-probability antecedents, $p(A) > 0$) and allows for properly managing zero-probability antecedents. The latter property is important, for example, to avoid counterintuitive inferences, like the following *paradox of the material conditional*:

C, therefore *if A, then C*,

which is logically valid (when the conditional is interpreted as a material one) but this argument form may have counterintuitive instantiations. Consider, for example, the following instantiation:

the weather is nice, therefore *if it is raining, then the weather is nice*,

which is an odd inference of course. The oddness of this inference is captured by *coherence-based probability logic*, which is about transmitting the uncertainty of the premises to the conclusion in a coherent way. Within coherence-based probability logic, the previous argument form is probabilistically

[6] Tentative interpretations of attack in modal logical terms have been considered in [22]. This semantics however does not generalize to attack principles with varying strength.

non-informative. That is, for all probability values of $p(C)$ (including 1), the tightest coherent bounds on the conclusion $p(C|A)$ coincide with the unit interval [0, 1] (for a proof see [66]). Since the unit interval does not restrict the degree of belief in the conclusion, this paradox is blocked in the coherence approach. However, this paradox arises within approaches which use the fraction definition of conditional probability, since in the particular case when $p(C) = 1$ (and where $p(A) > 0$ must be assumed to avoid fractions over zero), the conclusion $p(C|A)$ is assessed with the point-value 1, which is of course highly informative ([66]). Not only because of such technical virtues but also because of its empirically confirmed psychological plausibility (see, e.g., [57,64,67–69,72]), we use the coherence approach to probability in our paper.

Concretely, we suggest to read "F attacks A" as the assertion that it is likely that A does not hold, given that F holds. More precisely, we suggest an interpretation of $F \longrightarrow A$ as $p(\neg A|F) \geqslant t$, which is *parameterized* for some threshold $0.5 < t \leqslant 1$. We note that $p(\neg A|F) \geqslant t$ is equivalent to $p(A|F) < t$. The latter formulation is simpler, but since we are interested in measuring the strength of attack, we prefer the former formulation, which provides a lower bound on the strength of attack. Throughout the rest of the paper, we assume that F is not a logical contradiction (i.e., F is not equivalent to $\bot$, where $\bot$ denotes the truth constant *falsum*). This assumption is not only intuitively plausible (because assuming $\bot$ to be true does not make sense) but also technically important for us, since although zero-probability antecedents are allowed within the framework of coherence, $p(A|\bot)$ is undefined. Note that coherence requires that $p(A)$ must be equal to zero if A is a logical contradiction (since $\bot$ cannot be true, in betting terms, you can never win when you bet on the truth of $\bot$), while the reverse does not hold: if $P(A) = 0$, this does not mean that A is a logical contradiction, i.e., A could be contingent. Approaches to probability, where the values 0 and 1 are reserved for contradiction and tautology, respectively, are sometimes called "regular". The coherence approach is more general than approaches using regular probabilities, as 0 and 1 can also be assigned to contingent events.

We emphasize that the suggested interpretation of $F \longrightarrow A$ does not determine a specific interpretation of the attack relation between arguments of Λ itself. In particular, attack between *arguments* does not have to be defined in terms of probability. Only the relation between corresponding *claims of arguments* is interpreted probabilistically.

Meteorological Example 4. Once more, let R stand for "It will rain tomorrow" and S for "It will be sunny tomorrow" and suppose that R occurs as the claim of an argument that attacks an argument with claim S (i.e. $R \longrightarrow S$) in the underlying AF $\mathcal{M}$. To apply our interpretation, we first have to select a threshold value t such that $0.5 < t \leqslant 1$. Then, according to the interpretation, $\mathcal{M}$ contains information that indicates that the probability that it will not be sunny tomorrow, under the condition that it will actually rain tomorrow, is higher than t ($p(\neg S|R) > t$).

Note that our probabilistic interpretation operates on the semi-abstract (or claim-centric) level that shifts attention from individual attacks between arguments to the mere existence of attacks between arguments featuring certain claims. In this manner, adopting our interpretation amounts to imposing certain rationality constraints about the overall coherence of a given AF on the level of considered claims. So far, we only demand that each claim should always be possibly true according to some interpretation, i.e. it should not be self-contradictory. Further constraints of this kind will arise from the corresponding interpretation of our logical attack principles.

Translating the attack principles that refer to conjunction, disjunction, and negation according to the suggested interpretation is straightforward. The following probabilistic constraints correspond to the principles (**C.**$\wedge$), (**C.**$\vee$), and (**C.**$\neg$):

$(\mathbf{C.}\wedge)^t_p$ If $p(\neg A|F) \geqslant t$ or $p(\neg B|F) \geqslant t$, then $p(\neg(A \wedge B)|F) \geqslant t$.
$(\mathbf{C.}\vee)^t_p$ If $p(\neg(A \vee B)|F) \geqslant t$, then $p(\neg A|F) \geqslant t$ and $p(\neg B|F) \geqslant t$.
$(\mathbf{C.}\neg)^t_p$ If $p(\neg A|F) \geqslant t$, then $p(\neg\neg A|F) = p(A|F) < t$.

Analogously, the inverse principles translate as follows:

$(\mathbf{C.}\wedge)\text{'}^t_p$ If $p(\neg(A \wedge B)|F) \geqslant t$ then $p(\neg A|F) \geqslant t$ or $p(\neg B|F) \geqslant t$.
$(\mathbf{C.}\vee)\text{'}^t_p$ If $p(\neg A|F) \geqslant t$ and $p(\neg B|F) \geqslant t$ then $p(\neg(A \vee B)|F) \geqslant t$.
$(\mathbf{C.}\neg)\text{'}^t_p$ If $p(\neg A|F) < t$ then $p(\neg\neg A|F) = p(A|F) \geqslant t$.

If $p(A)$ and $p(B)$ are stochastically independent, then $p(A \wedge B) = p(A) \cdot p(B)$. However, if stochastic independence cannot be presupposed, the probability of the conjunction of A and B is given by the lower and upper Fréchet bounds, $\max\{0, p(A) + p(B) - 1\}$ and $\min\{p(A), p(B)\}$, respectively. The same bounds hold for corresponding conditional probabilities. The corresponding rule coincides with the probabilistic version of the (**And**) Rule of System P, which is among the most prominent systems of nonmonotonic reasoning [58]. It is proven to be coherent in [41]:

$(\mathbf{And})_p$ From $p(A|F) = x$ and $p(B|F) = y$ infer $\max(0, x + y - 1) \leqslant p(A \wedge B|F) \leqslant \min(x, y)$.

These coherent lower and upper bounds on the conclusion are the tightest or best-possible ones, which means that violating at least one of these bounds would make the probabilistic assessment incoherent.

Definition 3. We say that an attack principle ($\mathbf{C.}\circ$) *holds for a threshold* t (in the sense of coherence-based probability logic) if and only if the corresponding probability constraints $(\mathbf{C.}\circ)^t_p$, parameterized with respect to t, can be proven (within coherence-based probability logic) to be coherent according to Definition 1.

Proposition 1. ($\mathbf{C.}\wedge$), ($\mathbf{C.}\vee$), *and* ($\mathbf{C.}\neg$) *hold for every threshold* $t > .5$ *(cf. Definition 3). However,* ($\mathbf{C.}\wedge$)' *and* ($\mathbf{C.}\neg$)' *do not hold in this sense, for any threshold* $t > .5$. ($\mathbf{C.}\vee$)' *holds for* $t = 1$ *but does not hold for any* $0.5 < t < 1$.

Proof. For proving that principle ($\mathbf{C.}\wedge$) holds for every t in our probabilistic semantics, we recall that we use classical logic, hence $\neg(A \wedge B) \equiv \neg A \vee \neg B$. Therefore, $p(\neg(A \wedge B)|F) = p(\neg A \vee \neg B|F)$. Since the (conditional) probability of a disjunction is greater than or equal to the (conditional) probability of each of its disjuncts, $(\mathbf{C.}\wedge)^t_p$ is coherent.

We recall that, since $p(\neg(A \vee B)|F) = p(\neg A \wedge \neg B|F)$, the coherence of $(\mathbf{C.}\vee)^t_p$ is justified by the upper Fréchet bound, i.e., both $p(\neg A|F)$ and $p(\neg B|F)$ must be at least equal to $p(\neg A \wedge \neg B|F)$. Hence, ($\mathbf{C.}\vee$) holds for every t.

Since $p(\neg A|F) = 1 - p(A|F)$, $(\mathbf{C.}\neg)^t_p$ is coherent.

To see that ($\mathbf{C.}\wedge$)' does not hold for any $t > .5$, consider $p(\neg A|F) = p(\neg B|F) = .5$ and assume that $\neg A$ is equivalent to B. Then, $p(\neg A \vee \neg B|F) = p(\neg A|F) + p(\neg B|F) = 1$. Since $(\neg A \vee \neg B) \equiv \neg(A \wedge B)$, $(\mathbf{C.}\wedge)\text{'}^t_p$ is not satisfied for any threshold $t > .5$.

For proving that ($\mathbf{C.}\vee$)' does not hold for $t < 1$, assume, for example, that $p(\neg A|F) = p(\neg B|F) = .5$. Then, $p(\neg(A \vee B)|F)$ could still be strictly less than .5 (even equal to zero), since $(\mathbf{C.}\vee)\text{'}^t_p$ is an instance of $(\mathbf{And})_p$ (recall that $\neg(A \vee B) \equiv (\neg A \wedge \neg B)$). Hence, $(\mathbf{C.}\vee)\text{'}^t_p$ is not satisfied in general. In the particular case when $t = 1$, ($\mathbf{C.}\vee$)' holds since: if $p(\neg A|F) = 1$ and $p(\neg B|F) = 1$, then $(\mathbf{C.}\vee)\text{'}^t_p$ is coherent since $p(\neg(A \vee B)|F) = p(\neg A \wedge \neg B|F) = 1$, which is an instance of $(\mathbf{And})_p$.

$(\mathbf{C}.\neg)'^{t}_{p}$ is not coherent for any $t > .5$: if $p(\neg A|F) = p(A|F) = .5$, then $p(\neg A|F) < t$, but also $p(A|F) < t$. □

Meteorological Example 5. Recall that according to our probabilistic interpretation, $R \longrightarrow S$ expresses that, under the assumption that it will rain tomorrow, it is more likely that it will not be sunny tomorrow than that it will be sunny. Without imposing any restrictions on the attack relation, it may be the case that also $R \longrightarrow \neg S$ holds in the underlying AF $\mathcal{M}$. However, no coherent probabilities can be assigned to the conditional events $S|R$ and $\neg S|R$ according to $\mathcal{M}$, such that $p(S|R)$ and $p(\neg S|R)$ are both above $t > .5$. The fact that $(\mathbf{C}.\neg)^{t}_{p}$ is coherent (Proposition 1) entails that this cannot happen if $\mathcal{M}$ satisfies the logical attack principle (**C.**¬).

We now turn to arguments featuring conditionals as claims. As we use classical logic, $A \supset B$ is equivalent to $\neg A \vee B$. The corresponding translations of (**C.** ⊃) and (**C.** ⊃)**'** are as follows:

$(\mathbf{C}.\supset)^{t}_{p}$ If $p(\neg(A \supset B)|F) \geqslant t$ then $p(\neg B|F) \geqslant t$.
$(\mathbf{C}.\supset)'^{t}_{p}$ If $p(\neg B|F) \geqslant t$ then $p(\neg(A \supset B)|F) \geqslant t$.

Proposition 2. (**C.** ⊃) *holds in the sense of Definition 3, but* (**C.** ⊃)**'** *does not hold in this sense.*

Proof. As we use classical logic $A \supset B$ is equivalent to $\neg A \vee B$. Hence $(\mathbf{C}.\supset)^{t}_{p}$ turns into instance of $(\mathbf{C}.\vee)^{t}_{p}$. Therefore, by Proposition 1, $(\mathbf{C}.\supset)^{t}_{p}$ is coherent. Concerning $(\mathbf{C}.\supset)'^{t}_{p}$, note that $\neg(A \supset B) \equiv (A \wedge \neg B)$. Hence, $(\mathbf{C}.\supset)'^{t}_{p}$ is not coherent since $p(\neg B|F)$ may be strictly higher than $p(A \wedge \neg B|F)$. □

We note that interpreting attack principles involving the implication connective is delicate in general, since it is widely agreed that the natural language conditional ('if ..., then ...') should not be identified with classical (truth-functional) implication. Actually, as argued, e.g., in [42,66], coherence-based conditional probability itself provides a sound and robust semantics for the conditional. Moreover, the coherence approach turned out to be very useful for modeling argument strength [65,70]. Following this insight would force us to use degrees of beliefs in nested conditionals (e.g., in terms of previsions in conditional random quantities; see, e.g., [44,77–79]) to interpret principles like (**C.** ⊃). While this is an interesting topic for future research, here we only want to check how our probability-based interpretation of the attack relation classifies (**C.** ⊃) and (**C.** ⊃)**'** when classical logic is assumed. Therefore, we have chosen to use the material conditional interpretation of conditionals in our analysis.

In [22] also logically contradictory claims are considered by formulating the following corresponding attack principle:

$(\mathbf{C}.\bot)$ $F \longrightarrow \bot$.

In other words, assuming that ⊥ occurs as a claim in the underlying AF, (**C.**⊥) stipulates that for every argument featuring claim F there exists an argument claiming F that attacks an argument featuring ⊥ as its claim. In everyday life argumentation a contradictory claim is not attacked by arbitrary arguments, rather by simply pointing out that there is a contradiction. This is a pragmatic aspect of argumentation. However, here, we are solely concerned with semantic relations among claims.

The principle (**C.**⊥) is probabilistically interpreted by

$(\mathbf{C}.\bot)_{p}$ $p(\top|F) = 1$,

where $\top$ denotes the truth constant *verum*. Since coherence requires that $p(\neg\bot|F) = p(\top|F) = 1$, $(\mathbf{C}.\bot)_p$ is satisfied. However, note that we cannot interpret any principles that involve contradictory claims of attacking arguments, since the corresponding conditional probability must remain undefined.[7]

5. Quantitative attack principles and their semantics

So far, we have only discussed qualitative attack principles, i.e., principles that only care for the presence or absence of an attack between (claims of) given arguments. However it is natural to refine such an analysis by considering *weights* or *varying strength* of attacks. Various suggestions regarding weighted AFs can be found in the literature on argumentation in AI, see, e.g., [1,3,5,9,17,24,26,32]. But, to our best knowledge, there is no investigation yet of rationality postulates that systematically relates weights of explicit and implicit attacks to the *logical form* of the involved claims of arguments. However, see Section 8 for some remarks on, at least vaguely, related work.

A first step in that direction has been attempted in [23], where the principles introduced in [22] are generalized to the context of weighted AFs. The aim of [23] is to explore under which assumptions one can characterize various t-norm-based fuzzy logics in terms of 'weighted attack principles'. As expected, it turns out that some of the principles that are needed to recover a truth-functional (fuzzy) semantics are implausible from an intuitive, argumentation-based point of view. In any case, the situation, once more, calls for a systematic interpretation of the relevant principles, that enables one to formally judge their respective plausibility. Fortunately, the probabilistic interpretation of the qualitative attack principles, developed in Section 4, generalizes in a very direct and natural manner to the quantitative scenario.

Rather than just distinguishing between $F \longrightarrow A$ and $F \not\longrightarrow A$ ("F attacks / does not attack A"), we will use $F \overset{w}{\longrightarrow} A$ to denote that F attacks A with the weight (or to the degree) w. Let us stress again that "attack", here, is a relation between propositions and not between arguments. In the literature, there are various suggestions for generalizing ordinary AFs to weighted AFs (or systems), where real numbers attached to attacks between arguments are intended to represent degrees of strength of such attacks (see, in particular, [32]). In analogy to the qualitative scenario of Sections 3 and 4, one may understand $F \overset{w}{\longrightarrow} A$ to refer to an underlying weighted AF in some specific manner. For example, one might want to identify w with either the average or with the minimum of weights of all attacks of arguments with claim F on arguments with claim A and set $w = 0$ if no corresponding attack exists. However, since we are only interested in rationality constraints arising for logically complex antagonistic claims, we will treat weights of attacks between propositions as primitive, here. These weights are understood to be normalized, with 1 being the maximal weight of any attack, whereas $F \overset{0}{\longrightarrow} A$ means that F in fact does not attack the claim A at all. Note that this stipulation entails that the qualitative scenario discussed in Sections 3 and 4 amounts to an instance of the weighted case, where the only possible weights are 0 and 1. We deliberately refrain from prescribing specific, context-dependent methods for determining concrete weights of attacks, since we are interested in principles that do not depend on the specific content of the involved statements, but only on their logical form.

Meteorological Example 6. Continuing the meteorological example of the previous sections, we now imagine that the attacks between the various arguments regarding the weather forecast are weighted. We

[7] For a conditional sentence in ordinary language, it feels odd to assume that its antecedent is true, if it is a contradiction. We may, however, very well say, for example, that the probability of heads in a second toss is .5, if the coin lands on its edge in the first toss (under common assumptions about fair coins like $p(\text{heads}) = p(\text{tails}) = .5$ and $p(\text{coin lands on its edge}) = 0$).

deliberately ignore *how* the individual weights on attacks are determined, but assume that those weights a normalized, such that all attached weights are in $(0, 1]$. We call this weighted AF $\mathcal{M}_w$. Again, we consider the following statements that appear as claims of arguments in $\mathcal{M}_w$.

R**:** "It will rain tomorrow."
S**:** "It will be sunny tomorrow."
W**:** "It will be warm tomorrow."

As indicated above, we have to fix some mechanism for mapping weights between attacks into weights between claims. For sake of concreteness, we use the *supremum* over all weights of attacks of corresponding claims, if there is any. If there is no such attack we set the weight of attack between the corresponding claims to 0. This means that, e.g., $R \overset{0.8}{\longrightarrow} S$ is obtained by inspecting the set of all pairs (ϕ, α) of arguments in $\mathcal{M}_w$, where ϕ claims R and α claims S and ϕ attacks α. In our example 0.8 is the supremum over all weights of attacks of this type. On the other hand, we might find that $R \overset{0}{\longrightarrow} W$ in $\mathcal{M}_w$. This means that none of the arguments in $\mathcal{M}_w$ that claim that it will be sunny tomorrow attacks any argument in $\mathcal{M}_w$ claiming that it will be warm tomorrow.

An attractive feature of the probabilistic approach taken here is the fact that it immediately leads to a quantitative refinement of the qualitative case: interpreting attacks in terms of coherent conditional probabilities suggests to directly attach weights, instead of using thresholds to judge whether a given statement attacks another one. As pointed out in [23], there are several non-equivalent ways in which the qualitative attack principles reviewed in Section 3 can be generalized to 'weighted attack principles'. The most straightforward generalization of principle (**C.**$\wedge$) to weighted attacks is arguably the following:

If $F \overset{x}{\longrightarrow} A$ and $F \overset{y}{\longrightarrow} B$, then $F \overset{z}{\longrightarrow} A \wedge B$, where $z \geqslant \max(x, y)$.

Actually, since we also consider attacks of weight 0 (interpreted as 'no attack'), we may assume without loss of generality that there is a weighted attack between any pair of formulas. This means that the above principle can be reformulated as a constraint on the corresponding weights as follows:

($\mathbf{G}^w_{\geqslant}$**.**$\wedge$) If $F \overset{x}{\longrightarrow} A$, $F \overset{y}{\longrightarrow} B$, and $F \overset{z}{\longrightarrow} A \wedge B$, then $z \geqslant \max(x, y)$.

Meteorological Example 7. Continuing Example 6, consider $R \overset{0.8}{\longrightarrow} S$ and $R \overset{0.6}{\longrightarrow} W$. Recall that in our example the indicated weights refer to the maximal weights of attacks of arguments claiming that it will rain tomorrow on arguments claiming that it will be sunny tomorrow or on arguments claiming that it will be warm tomorrow, respectively. If the underlying weighted AF $\mathcal{M}_w$ satisfies the attack principle ($\mathbf{G}^w_{\geqslant}$**.**$\wedge$), then among all arguments that claim that it will be sunny *and* warm tomorrow in $\mathcal{M}_w$, at least one is attacked with weight $w \geqslant 0.8$ by some argument claiming that it will rain tomorrow.

Alternative weighted attack principles for conjunction, formulated in the same manner, are:

($\mathbf{Ł}^w_{\geqslant}$**.**$\wedge$) If $F \overset{x}{\longrightarrow} A$, $F \overset{y}{\longrightarrow} B$, and $F \overset{z}{\longrightarrow} A \wedge B$, then $z \geqslant \min(1, x + y)$.
($\mathbf{P}^w_{\geqslant}$**.**$\wedge$) If $F \overset{x}{\longrightarrow} A$, $F \overset{y}{\longrightarrow} B$, and $F \overset{z}{\longrightarrow} A \wedge B$, then $z \geqslant x + y - xy$.

As the labels indicate, these principles are essential for obtaining an argumentation-based semantics for Gödel logic **G**, Łukasiewicz logic **Ł**, and Product logic **P**, respectively. These three logics are the most

fundamental t-norm-based fuzzy logics, since any fuzzy logic based on a continuous t-norm as truth-function for conjunction can be represented in terms of **G**, **Ł**, and **P** [18,51]. Moreover the subscript '$\geqslant$' attached to these letters indicates that we formulate here upper bounds on the weight of attacks of conjunctive claims (in terms of weights of attacks on conjuncts). In fact, also principles expressing matching lower bounds are needed to characterize the three mentioned t-norm-based fuzzy logics. Correspondingly, we use $(\mathbf{G}^{w}_{\leqslant}.\wedge)$, $(\mathbf{Ł}^{w}_{\leqslant}.\wedge)$, and $(\mathbf{P}^{w}_{\leqslant}.\wedge)$ to refer to the principles that arise by just replacing '$\geqslant$' by '$\leqslant$' in the respective constraint.

As already indicated, in contrast to the qualitative case of Section 4, we do not have to involve threshold values in interpreting a weighted attack relation between claims, but simply identify the weight with which F attacks A with the conditional probability that A does not hold, given that F holds. More formally, our probabilistic semantics interprets $F \overset{w}{\longrightarrow} A$ by $p(\neg A|F) = w$. (Remember that this is only viable if we exclude the possibility that F is a logical contradiction; although, we allow for the possibility that $p(F) = 0$.) Once more, we point out that interpreting weights between *claims* of arguments as probabilities does not mean that we have to interpret also weights of the underlying attacks between *arguments* probabilistically. The suggested semantics operates on the semi-abstract level that deliberately ignores the fully instantiated level of attacks between concrete arguments, which may well depend on the support part of arguments and not just their claims.

According to the probabilistic semantics, the above versions of weighted attack principles translate into the following statements:

$(\mathbf{G}^{w}_{\geqslant}.\wedge)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \wedge B)|F) \geqslant \max(x, y)$.
$(\mathbf{Ł}^{w}_{\geqslant}.\wedge)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \wedge B)|F) \geqslant \min(1, x + y)$.
$(\mathbf{P}^{w}_{\geqslant}.\wedge)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \wedge B)|F) \geqslant x + y - xy$.
$(\mathbf{G}^{w}_{\leqslant}.\wedge)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \wedge B)|F) \leqslant \max(x, y)$.
$(\mathbf{Ł}^{w}_{\leqslant}.\wedge)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \wedge B)|F) \leqslant \min(1, x + y)$.
$(\mathbf{P}^{w}_{\leqslant}.\wedge)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \wedge B)|F) \leqslant x + y - xy$.

Meteorological Example 8. Let us apply the probabilistic interpretation of weighted attacks between claims to Example 7. This means that the underlying weighted AF $\mathcal{M}_w$ can be understood to contain information indicating that the probability that it will not be sunny tomorrow, given that it will rain tomorrow, is 0.8 ($p(\neg S|R) = 0.8$). Similarly, there is information indicating that the probability that it will not be warm tomorrow, under the condition that it will rain tomorrow is 0.6 ($p(\neg W|R) = 0.6$). The probabilistic interpretation $(\mathbf{G}^{w}_{\geqslant}.\wedge)_p$ of the attack principle $(\mathbf{G}^{w}_{\geqslant}.\wedge)$ stipulates that $\mathcal{M}_w$ contains information according to which the probability that it will not be sunny as well as warm tomorrow, given that it will rain tomorrow, is at least 0.8 ($p(\neg(S \wedge W)|R) \geqslant 0.8$).

We now investigate which of the various possible weighted attack principles for conjunction should indeed be adopted as rationality principles constraining the underlying AFs, if we follow the interpretation of weights of attacks between claims as coherent conditional probabilities. We obtain the following corresponding classification.

Proposition 3. *The principles* $(\mathbf{G}^{w}_{\geqslant}.\wedge)$ *and* $(\mathbf{Ł}^{w}_{\leqslant}.\wedge)$ *hold (i.e., the constraints are coherent in the sense of Definition 1). However, the principles* $(\mathbf{Ł}^{w}_{\geqslant}.\wedge)$, $(\mathbf{P}^{w}_{\geqslant}.\wedge)$, $(\mathbf{G}^{w}_{\leqslant}.\wedge)$, *and* $(\mathbf{P}^{w}_{\leqslant}.\wedge)$ *do not hold for all coherent probability assessments.*

Proof. Remember that we assume that all involved propositions are classical. Hence, $\neg(A \wedge B)$ is equivalent to $\neg A \vee \neg B$. Since $p(\neg A|F) \leqslant p(\neg A \vee \neg B|F)$ and $p(\neg B|F) \leqslant p(\neg A \vee \neg B|F)$, $(\mathbf{G}^w_{\geqslant}.\wedge)_p$ is coherent. Concerning $(\mathbf{Ł}^w_{\leqslant}.\wedge)_p$, let $p(\neg A|F) = x$ and $p(\neg B|F) = y$. The law of additivity for conditional probability requires that $p(\neg A \vee \neg B|F) = x + y - p(\neg A \wedge \neg B|F)$, which is always smaller than or equal to $\min(1, x + y)$. Hence, $(\mathbf{Ł}^w_{\leqslant}.\wedge)_p$ is satisfied.

The corresponding probabilistic constraints for the four other principles can be violated:

$(\mathbf{Ł}^w_{\geqslant}.\wedge)_p$, $(\mathbf{P}^w_{\geqslant}.\wedge)_p$**:** Let $A = B$ and $p(\neg A|F) = p(\neg B|F) = 0.5$. Then $p(\neg(A \wedge B)|F) = p(\neg(A \wedge A)|F) = p(\neg A|F) = 0.5$, which is strictly smaller than $\min(1, 0.5 + 0.5) = 1$, but also strictly smaller than $0.5 + 0.5 - 0.5^2 = 0.75$.

$(\mathbf{G}^w_{\leqslant}.\wedge)_p$, $(\mathbf{P}^w_{\leqslant}.\wedge)_p$**:** Let $A = \neg B$ and $p(\neg A|F) = p(\neg B|F) = 0.5$. Then $p(\neg(A \wedge B)|F) = p(\neg(A \wedge \neg A)|F) = p(\neg\bot|F) = p(\top|F) = 1$, which is strictly larger than $\max(0.5, 0.5) = 0.5$ and strictly larger than $0.5 + 0.5 - 0.5^2 = 0.75$. □

Although principles $(\mathbf{Ł}^w_{\geqslant}.\wedge)$, $(\mathbf{P}^w_{\geqslant}.\wedge)$, $(\mathbf{G}^w_{\leqslant}.\wedge)$, and $(\mathbf{P}^w_{\leqslant}.\wedge)$ do not hold generally under coherence, the corresponding conditions $(\mathbf{Ł}^w_{\geqslant}.\wedge)_p$, $(\mathbf{P}^w_{\geqslant}.\wedge)_p$, $(\mathbf{G}^w_{\leqslant}.\wedge)_p$, and $(\mathbf{P}^w_{\leqslant}.\wedge)_p$, respectively, may hold for *particular* probability assignments. Consider, for example, the following propositions:

Proposition 4. *Under the assumption that $p(A|F)$ and $p(B|F)$ are independent, $(\mathbf{P}^w_{\geqslant}.\wedge)_p$ and $(\mathbf{P}^w_{\leqslant}.\wedge)_p$ hold.*

Proof. If $\neg A$ and $\neg B$ are conditionally independent given F, the probability of the conjunction of $\neg A$ and $\neg B$ given F is the product of the probabilities of the conditional events $\neg A|F$ and $\neg B|F$: $p(\neg A \wedge \neg B|F) = p(\neg A|F) \cdot p(\neg B|F)$. Hence, $p(\neg A \vee \neg B|F) = p(\neg A|F) + p(\neg B|F) - p(\neg A|F) \cdot p(\neg B|F)$ and therefore $(\mathbf{P}^w_{\geqslant}.\wedge)_p$ and $(\mathbf{P}^w_{\leqslant}.\wedge)_p$ hold under this assumption. □

Proposition 5. *Under the assumption that $A \models B$ or $B \models A$, $(\mathbf{G}^w_{\leqslant}.\wedge)_p$ holds.*

Proof. If $A \models B$, then $\neg B \models \neg A$. Hence, $p(\neg A \vee \neg B|F) = p(\neg A|F)$. Recall that $p(\neg(A \wedge B)|F) = p(\neg A \vee \neg B|F)$. Therefore, $p(\neg(A \wedge B)|F) = p(\neg A|F) \leqslant \max(p(\neg A|F), p(\neg B|F))$. The case for $B \models A$ is analogous. □

Proposition 6. *Under the assumption that $\neg A \models B$ (or, equivalently, $\neg B \models A$), $(\mathbf{Ł}^w_{\geqslant}.\wedge)_p$ holds.*

Proof. Observe that $\neg A \models B$ entails $\neg A \wedge \neg B \equiv \bot$, which means that $\neg A$ and $\neg B$ represent disjoint events. Hence $p(\neg A \vee \neg B|F) = p(\neg A|F) + p(\neg B|F)$. Therefore, $p(\neg(A \wedge B)|F) = p(\neg A \vee \neg B|F) \geqslant \min(1, p(\neg A|F) + p(\neg B|F))$. □

From the just outlined evaluation for the attack principles involving conjunction, we now turn to attack principles involving disjunction, which are of course, dual to those for conjunction.

$(\mathbf{G}^w_{\geqslant}.\vee)$ If $F \xrightarrow{x} A$, $F \xrightarrow{y} B$, and $F \xrightarrow{z} A \vee B$, then $z \geqslant \min(x, y)$.

$(\mathbf{Ł}^w_{\geqslant}.\vee)$ If $F \xrightarrow{x} A$, $F \xrightarrow{y} B$, and $F \xrightarrow{z} A \vee B$, then $z \geqslant \max(0, x + y - 1)$.

$(\mathbf{P}^w_{\geqslant}.\vee)$ If $F \xrightarrow{x} A$, $F \xrightarrow{y} B$, and $F \xrightarrow{z} A \vee B$, then $z \geqslant xy$.

Likewise, we use $(\mathbf{G}^w_{\leqslant}.\vee)$, $(\mathbf{Ł}^w_{\leqslant}.\vee)$, and $(\mathbf{P}^w_{\leqslant}.\vee)$ to refer to the principles that arise by just replacing '$\geqslant$' by '$\leqslant$' in the respective constraint and we obtain the following proposition:

Proposition 7. *The principles* $(\mathbf{G}^w_{\leqslant}.\vee)$ *and* $(\mathbf{Ł}^w_{\geqslant}.\vee)$ *hold. However, the principles* $(\mathbf{Ł}^w_{\leqslant}.\vee)$, $(\mathbf{P}^w_{\leqslant}.\vee)$, $(\mathbf{G}^w_{\geqslant}.\vee)$, *and* $(\mathbf{P}^w_{\geqslant}.\vee)$ *do not hold for all coherent probability assessments.*

Proof. $(\mathbf{G}^w_{\leqslant}.\vee)$ and $(\mathbf{Ł}^w_{\geqslant}.\vee)$ hold, since $p(\neg(A \vee B)|F) = p(\neg A \wedge \neg B|F)$ and $(\mathbf{G}^w_{\leqslant}.\vee)_p$ and $(\mathbf{Ł}^w_{\geqslant}.\vee)_p$ are instantiations of $(\mathbf{And})_p$:

$(\mathbf{Ł}^w_{\geqslant}.\vee)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \vee B)|F) \geqslant \max(0, x + y - 1)$.
$(\mathbf{G}^w_{\leqslant}.\vee)_p$ If $p(\neg A|F) = x$ and $p(\neg B|F) = y$ then $p(\neg(A \vee B)|F) \leqslant \min(x, y)$.

The corresponding probabilistic constraints for the four other principles can be violated:

$(\mathbf{Ł}^w_{\leqslant}.\vee)_p$, $(\mathbf{P}^w_{\leqslant}.\vee)_p$: Let $A = B$ and $p(\neg A|F) = p(\neg B|F) = 0.5$. Then $p(\neg(A \vee B)|F) = p(\neg(A \vee A)|F) = p(\neg A|F) = 0.5$, which is strictly greater than $\max(0, 0.5 + 0.5 - 1) = 0$, but also strictly greater than $0.5^2 = 0.25$.

$(\mathbf{G}^w_{\geqslant}.\vee)_p$, $(\mathbf{P}^w_{\geqslant}.\vee)_p$: Let $A = \neg B$ and $p(\neg A|F) = p(\neg B|F) = 0.5$. Then $p(\neg(A \vee B)|F) = p(\neg(A \vee \neg A)|F) = p(\neg\top|F) = p(\bot|F) = 0$, which is strictly smaller than $\min(0.5, 0.5) = 0.5$ and strictly smaller than $0.5^2 = 0.25$. □

Analogous results can be obtained for principles involving conditionals, since the material conditional $A \supset B$ is logically equivalent to the disjunction $\neg A \vee B$. For example, for Gödel logic we obtain the following two principles:

$(\mathbf{G}^w_{\leqslant}.\supset)$ If $F \xrightarrow{y} B$ and $F \xrightarrow{z} A \supset B$, then $z \leqslant y$.
$(\mathbf{G}^w_{\geqslant}.\supset)$ If $F \xrightarrow{y} B$ and $F \xrightarrow{z} A \supset B$, then $z \geqslant y$.

The principles $(\mathbf{G}^w_{\geqslant}.\supset)$ and $(\mathbf{G}^w_{\leqslant}.\supset)$ are interpreted, respectively, as follows:

$(\mathbf{G}^w_{\leqslant}.\supset)_p$ If $p(\neg B|F) = y$, then $p(\neg(A \supset B)|F) \leqslant y$.
$(\mathbf{G}^w_{\geqslant}.\supset)_p$ If $p(\neg B|F) = y$, then $p(\neg(A \supset B)|F) \geqslant y$.

Proposition 8. *The principle* $(\mathbf{G}^w_{\leqslant}.\supset)$ *holds, but* $(\mathbf{G}^w_{\geqslant}.\supset)_p$ *does not hold for all coherent probability assessments.*

Proof. $(\mathbf{G}^w_{\leqslant}.\supset)_p$ is satisfied, since $p(\neg(A \supset B)|F) = p(\neg(\neg A \vee B)|F) = p(A \wedge \neg B|F) \leqslant p(\neg B|F)$. Hence $(\mathbf{G}^w_{\leqslant}.\supset)$ is valid.

Let $A = \bot$ and $p(\neg B|F) = .5$. Then, $p(\neg(A \supset B)|F) = p(A \wedge \neg B|F) = 0$, which is less than .5. Therefore, $(\mathbf{G}^w_{\geqslant}.\supset)_p$ is not satisfied and hence $(\mathbf{G}^w_{\geqslant}.\supset)$ is not valid. □

Concerning negation, our semantics naturally suggests the following attack principle:

$(\mathbf{Ł}.\neg)$ If $F \xrightarrow{x} A$ and $F \xrightarrow{z} \neg A$, then $z = 1 - x$.

This is interpreted as follows:

$(\mathbf{Ł}.\neg)_p$ If $p(\neg A|F) = x$, then $p(\neg\neg A|F) = p(A|F) = 1 - x$.

The following proposition thus holds trivially:

Proposition 9. *The principle* (**Ł.¬**) *holds.*

We recall that according to the semantics of Łukasiewicz logic, if the truth value of A is x, then the truth value of $\neg A$ is $1 - x$, which coincides with the negation in probability theory as expressed in $(\mathbf{Ł.}\neg)_p$. However, negation in Gödel and Product logic is different: in both logics the truth value of $\neg A$ is 0 if the truth value of A is positive and 1 otherwise. Corresponding principles would not be justified within coherence-based probability semantics when classical negation is used.

Regarding falsum we obtain the following principle:

$(\mathbf{C}^w\mathbf{.}\bot)\quad F \xrightarrow{1} \bot,$

which is valid since it is interpreted by

$(\mathbf{C}^w\mathbf{.}\bot)_p\quad p(\neg\bot|F) = p(\top|F) = 1.$

Note that $(\mathbf{C}^w\mathbf{.}\bot)$ coincides with $(\mathbf{C.}\bot)$; consequently, also $(\mathbf{C}^w\mathbf{.}\bot)_p$ and $(\mathbf{C.}\bot)_p$ coincide. Moreover, for principle $(\mathbf{C}^w\mathbf{.}\bot)$ it is immaterial whether we refer to classical (**C**) or to many-valued logic (like **Ł**, **P**, or **G**; see [18,51]). Like in the corresponding qualitative case above, $(\mathbf{C}^w\mathbf{.}\bot)$ enforces a kind of homogeneity among the underlying arguments. If one wants to avoid this kind of homogeneity, one may consider an interpretation of weights in terms of belief functions [27,80], possibility measures [29], or ranking functions [81].

Regarding implication, one may of course extract corresponding principles from the above mentioned ones, under the stipulation that $A \supset B$ is understood, classically, as equivalent to $\neg A \vee B$. But, as already indicated, it would actually be more adequate to model (informal) implication not as a disjunction but as a proper conditional. This leads to the tricky and, as yet, only partially explored terrain of iterated conditional probabilities (for an approach within coherence, see, e.g., [43–46,77,78]); thus providing a challenging topic for future research.

6. Psychological experiment

In this section we present a first experiment which serves to explore empirically the psychological plausibility of the interpretation of the attack principles in our approach. Table 1 gives an overview on the investigated argument forms/formulas. Coherence-based probability logic received empirical support in recent years (e.g., [57,64,68,69,72]). However, principles governing the strength of attacks have not yet been investigated empirically (neither within nor outside the coherence framework; for an overview on empirical work on abstract argumentation see, e.g., [16]).

Participants. The sample consists of 139 computer science students who took part in the lecture *Formale Modellierung* at the TU Wien (Technical University of Vienna, 18 female, 116 male, and 5 who chose not to reveal their gender) with a mean age of 21.1 years ($SD = 3.2$). Only German native speakers were included in the data analysis. Seven participants were excluded from the analysis because of missing data in the target tasks. Most students were in their second semester and did not receive a thorough training in logic yet.

Table 2 shows that, on the average, the participants rated the overall task clearness[8] and difficulty[9] on an intermediate level ($M = 4.9$ and $M = 4.3$, respectively, on a rating scale out of 10). The intermediate

[8] *Sind die Aufgaben klar und verständlich formuliert?* (Are the tasks formulated clearly and comprehensively?)

[9] *Wie schwierig finden Sie die Aufgaben?* (How difficult are the tasks to you?)

Table 1

Task names/argument forms (or formulas) of the task sets of the three groups A, B, and C. Quantitative task types consist of correctness judgments (conditions A and B, see, e.g., Fig. 1) or of generations of strengths of attacks (condition C; see, e.g., Fig. 2). All three groups were also presented with qualtitative task types (with the three forced-choice options: wrong/correct/undetermined). "$A \overset{x}{\longrightarrow} B$" denotes "*A attacks by strength x the assertion B*", where x can be point- or interval-valued

Task name	Task/argument form	Task	Task type
Conjunction introduction	if $A \overset{x}{\longrightarrow} B$, then $A \overset{[x,1]}{\longrightarrow} (B \wedge C)$	B2,C4	quantitative
Conjunction elimination	if $A \overset{x}{\longrightarrow} (B \wedge C)$, then $A \overset{[0,x]}{\longrightarrow} B$	A1,C1	quantitative
Disjunction elimination	if $A \overset{x}{\longrightarrow} (B \vee C)$, then $A \overset{[x,1]}{\longrightarrow} B$	A2,C3	quantitative
Disjunction introduction	if $A \overset{x}{\longrightarrow} B$, then $A \overset{[0,x]}{\longrightarrow} (B \vee C)$	B3,C6	quantitative
Irrelevant premise	if $A \overset{x}{\longrightarrow} B$ and $C \models B$ then $A \overset{x}{\longrightarrow} B$	A3,C5	quantitative
(**Ł.¬**)	if $A \overset{x}{\longrightarrow} B$, then $A \overset{1-x}{\longrightarrow} \neg B$	A7,B1,B5,C2, C11	quantitative
(**Ł.¬**) variant	if $A \overset{x}{\longrightarrow} \neg B$, then $A \overset{1-x}{\longrightarrow} B$	A4,C7	quantitative
(**C.¬**)	if $A \longrightarrow B$, then $A \not\longrightarrow \neg B$	B11,C18	qualitative
(**C.¬**) variant	if $A \longrightarrow \neg B$, then $A \not\longrightarrow B$	B12,C19	qualitative
Attacked contradiction	$A \overset{1}{\longrightarrow} (B \wedge \neg B)$	B4,C9	qualitative
Attacked tautology	$A \overset{0}{\longrightarrow} (B \vee \neg B)$	B8,C15	qualitative
Negation attack	$\neg A \not\longrightarrow A$	B6,C12	qualitative
Negation attack′	$A \not\longrightarrow \neg A$	A5,C8	qualitative
Contradictory attack	not: $A \longrightarrow B$ and $A \longrightarrow \neg B$	B7,C14	qualitative
Reflexivity	$A \overset{0}{\longrightarrow} A$	A6,C10	qualitative
Contingent attack	$A \overset{[0,1]}{\longrightarrow} B$	A8,C13	quantitative
ProbToAttack	if $P(B\|A) = x$, then $A \overset{x}{\longrightarrow} \neg B$	A10,B9	quantitative
AttackToProb	if $A \overset{x}{\longrightarrow} B$, then $P(\neg B\|A) = x$	A9,B10	quantitative
AttackToProb′	if $A \overset{x}{\longrightarrow} B$, then $P(B\|A) = 1 - x$	C16	quantitative
ProbToAttack′	if $P(B\|A) = x$, then $A \overset{1-x}{\longrightarrow} B$	C17	quantitative

Table 2

Participants mean ratings on a scale coded from 1 to 10 (and standard deviations, *SD*) of the overall clearness of the tasks (10 = "clear"), their confidence in the correctness (10 = "confident") of their responses, the task difficulty (10 = "easy"), and whether they like to solve logical/mathematical tasks (10 = "like")

	clear	*SD*	conf.	*SD*	difficult	*SD*	like	*SD*
Task set A ($n_1 = 44$)	4.60	3.00	3.60	2.50	4.50	2.00	7.70	2.10
Task set B ($n_2 = 48$)	4.80	2.60	4.40	2.50	4.10	2.10	7.30	2.00
Task set C ($n_3 = 47$)	5.30	2.50	4.20	2.40	4.40	2.20	7.50	2.10

task difficulty ratings indicate that the tasks were neither perceived to be too easy nor too difficult: this is good, as extreme values on this scale could indicate a decreased motivation for solving the tasks. However, concerning the overall comprehensibility of the tasks, ratings closer to the maximum (10) would be preferable compared to the observed average close to 5, which hampers the interpretability of the data and restricts its conclusiveness. The intermediate comprehensibility of the tasks could be due to the implicit and explicit negations in the task material: psychological reasoning research indicates that negations are harder to process (see, e.g., [36]), which may thus negatively impact perceived task comprehensibility. More specifically, recall the distinction between *reasoning to* an interpretation and *reasoning from* a fixed interpretation [82]. This distinction refers to two general reasoning processes:

firstly, participants reason how to interpret the task; secondly, after fixing their interpretation, participants reason from their interpretation to the conclusion [82]. If the process of fixing the interpretation is hard (presumably because of negations), then the perceived task comprehensibility is lower compared to straightforward tasks. Moreover, processing problems during the reasoning to an interpretation may also explain why the participants were not highly confident in the correctness[10] of their solutions ($M = 4.1$ out of 10) even if in general they tend to like solving logical/mathematical problems[11] ($M = 7.5$ out of 10). Overall, we observed positive correlations between confidence in correctness and task difficulty ratings, $r(137) = .52$, $p < .001$, and between confidence in correctness and task comprehensibility ratings, $r(136) = .37$, $p < .001$: the higher the confidence in correctness ratings, the easier and more comprehensible the tasks were rated.[12] Interestingly, there was no statistically significant correlation between task comprehensibility and difficulty ratings, $r(136) = .16$, $p = 0.058$: whether the tasks were rated as comprehensible or not had no impact on how difficult the tasks were rated.

Method and materials. Each participant was given a A4 sheet of paper, containing an introduction on the first page and the target tasks on both pages. There were three between-participant conditions (to test inter-group differences), two with forced-choice (group A: $n_1 = 44$ and group B: $n_2 = 48$; see, e.g., Fig. 1) and one with an open choice response format (group C: $n_3 = 47$; see, e.g., Fig. 2). We hypothesised that the forced-choice tasks were easier compared to the open response format tasks, since judging attack strength candidates requires less cognitive effort and is hence considered to be easier compared to generating attack strengths. After showing that the degree of attack can be expressed on a scale form 0 to 10 and that claims can also be compounded (like **[A and B]**), the participants were

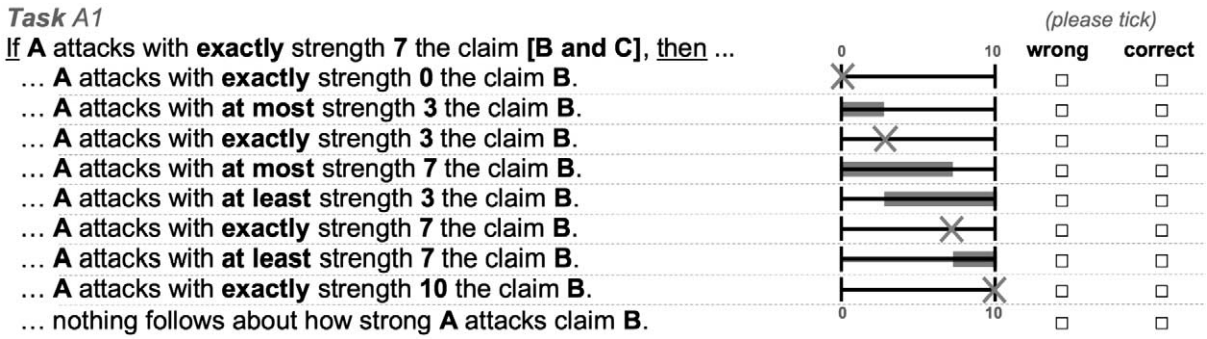

Fig. 1. Sample Task A1. Task type: quantitative, judgment of correctness task. Response format: forced-choice. For the corresponding argument form (Conjunction elimination) and the competence response, see Table 1.

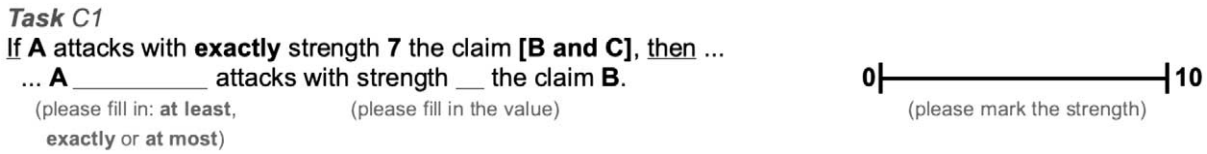

Fig. 2. Sample Task C1. Task type: quantitative, attack strength rating task. Response format: open. For the corresponding argument form (Conjunction elimination) and the competence response, see Table 1.

[10] *Wie sicher sind Sie, dass Ihre Lösungen stimmen?* (How sure are you that your solutions are correct?)
[11] *Lösen Sie gerne logische/mathematische Aufgaben?* (Do you like to solve logical/mathematical problems?)
[12] The degrees of freedom differ because one participant did not rate the task comprehensibility.

presented with tasks corresponding to the argument forms described in Table 1. The participants' task consists in evaluating possible consequents of the respective conditional, either in terms of *judging the correctness* of presented consequent candidates (groups A and B, forced-choice format, illustrated in Fig. 1) or in terms of *generating* strengths (group C, open response format, illustrated in Fig. 2). The Conjunction elimination tasks A1 and C1, for example, present the antecedent of the conditional "If **A** attacks with **exactly** the strength **7** the claim **[B and C]**, then ..." and differ in the way how the participants complete the conditional's consequent (compare Figs 1 and 2).

For the sake of simplicity, we omit references to underlying sets of arguments featuring such claims. Therefore, we assume that attacks can be viewed as directly relating claims rather than requiring reference to the possible complex underlying arguments. Moreover, we are interested in how logical form impacts on reasoning and not how contexts may influence the participants' interpretation of the attack principles. Therefore we have chosen abstract task materials. Since the sample consists of computer science students we did not expect *a priori* problems of understanding the (semi-)formal character of the tasks. Moreover, concrete task materials derived from everyday life examples may yield belief biases, i.e., that people may ignore the logical form and draw their inferences based on what they think is factually true (see, e.g., [37]).

For communicating quantitative degrees of attack we decided to use point-values and intervals constructed from the numbers 0, 3, 7, and 10 only. The reason for this restriction is to keep the task simple for the participants while still allowing for investigating quantitative attack principles. Tasks A9, B10, and C16 requested participants to respond in terms of probability judgments. Consequently, the values of the attack strength candidates were replaced by probabilities and the labels "0" and "10" in the illustrative scales were replaced by "$P(B|A) = 0$" and "$P(B|A) = 1$", respectively.

In the tasks of conditions A and B, nine consequent candidates were presented, which completed the conditional. Eight consequents were of the form "... attacks **A** with [M] with the strength [S] the claim **B**", where "[M]" indicates a precise value ("exactly"), a lower ("at least"), or an upper bound ("at most") on the strength [S]. [S] was either 0, 3, 7, or 10. All possible point and interval options were formulated in ascending order (see Table 3 for the attack strength options we used). Except for the interval [0, 10] we used "nothing follows about how strong ... attacks ...", as the ninth response option. The participants were asked to tick for each of the nine items whether the according sentence is wrong ("*falsch*") or correct ("*richtig*"). In the strength generation condition C, the participants were instructed to fill in "exactly", "at least", or "at most", the value of the strength, and additionally had to mark the strength of attack (either as a point value or as an interval) on a scale as introduced at the beginning of the instructions.

All participants were presented with quantitative and with qualitative task types (see Table 1). The quantitative task types were of the kind we just described. In the qualitative task types the participants were asked to choose one among the three options "wrong", "correct", or "undetermined" (*unbestimmt*) by ticking a corresponding box. The qualitative Negation attack′ task A5, for instance, asked the participants whether the following assertion is in principle wrong, correct or undetermined (*Ist folgende Behauptung grundsätzlich falsch, richtig oder unbestimmt?*): **It is not the case that: A** attacks **not-A**.

Procedure. The experiment took place during the last part of the first lecture of the course (*Formale Modellierung*). The students were informed that the experiment aims to investigate systematic relationships between logical form and argumentation, that their participation makes an important contribution to research, that the experiment is (not about testing skills but) about finding out how people deal with certain argument forms, that participation is voluntary, and that anonymity is guaranteed. Moreover, to foster independent processing of the tasks, we informed participants that there are different versions of

task sets. We also told the participants that they should first read carefully the introductions and then the tasks. We stressed that the individual claims differ: sometimes only in detail. The participants were asked to think carefully and to take as much time as they need for answering the questions. Then we made the importance of answering the questions independently explicit by also explaining that it would be very unfavorable for the statistical analysis and would distort the data if the participants influence each other during the experiment. We expected that not everyone will finish the filling in process at the same time. Thus, in order to further prevent conversations during the experiment and to keep the noise in the lecture hall down, we asked the participants to keep quiet and remain seated until the sheets are collected.

Then we continued with administering the task pages. The three conditions were administered in a systematically alternated way to reduce the chance of plagiarized responses.

Results and discussion. The main results are presented in Tables 3–7. First, we observe that most people are unaware of the best possible (or tightest) coherent bounds (marked in ***italics and bold***). Responses which are within the best possible coherent bounds are of course also coherent, like in the Conjunction elimination task A1 where 45% of the participants judged that "precisely 7" is correct. In this task, 43% judged that the interval "at most 7" is correct, which corresponds to the best possible coherent interval. The response patterns in the corresponding strength generation task C1 were analogous: most participants responded with a precise value equal to the coherent upper bound. Thereby, the participants neglect that the value zero is the best possible coherent lower bound. Second, we observe that compared to direct

Table 3

Quantitative task types: percentages of "correct" responses concerning the point valued/interval attack strength and nf (= "nothing follows") answer options in condition A ($n_1 = 44$). As task A9 asked for probabilities, these responses were rescaled to fit the table. Coherent responses are in *italics*. Best possible/tightest coherent response options are also ***bold*** (for the response options see Fig. 1; for the predictions see Table 1)

Task	Attack strength option								nf
	0	[0, 3]	3	[0, 7]	[3, 10]	7	[7, 10]	10	
A1	*0.00*	*0.00*	*0.00*	***43.18***	18.18	*45.45*	18.18	0.00	31.82
A2	0.00	0.00	0.00	63.64	6.82	*25.00*	***9.09***	*0.00*	34.09
A3	0.00	2.27	0.00	25.00	18.18	***93.18***	27.27	0.00	4.55
A4	20.45	18.18	***18.18***	11.36	2.27	2.27	0.00	0.00	59.09
A7	15.91	22.73	***20.45***	13.64	6.82	9.09	0.00	0.00	52.27
A8	*6.82*	*4.55*	*4.55*	*6.82*	*4.55*	*4.55*	*4.55*	*4.55*	***88.64***
A9	2.27	13.64	22.73	2.27	9.09	***13.64***	6.82	4.55	56.82
A10	4.55	4.55	***13.64***	2.27	9.09	11.36	11.36	2.27	63.64

Table 4

Quantitative task types: percentages of "correct" responses in condition B ($n_2 = 48$). As task B10 asked for probabilities, these responses were rescaled to fit the table. See also caption of Table 3

Task	Attack strength option								nf
	0	[0, 3]	3	[0, 7]	[3, 10]	7	[7, 10]	10	
B1	8.33	31.25	***29.17***	2.08	4.17	2.08	0.00	0.00	43.75
B2	2.08	4.17	2.08	22.92	18.75	*16.67*	***20.83***	*0.00*	39.58
B3	*2.08*	*4.17*	*2.08*	***27.08***	18.75	*25.00*	33.33	4.17	27.08
B5	8.33	31.25	***29.17***	0.00	4.17	0.00	2.08	4.17	45.83
B9	4.17	14.58	16.67	8.33	0.00	***2.08***	4.17	0.00	62.50
B10	2.08	12.50	14.58	8.33	4.17	***20.83***	4.17	0.00	47.92

Table 5

Qualitative task types: percentages of responses in conditions A ($n_1 = 44$) and B ($n_2 = 48$). (Best possible) coherent response options are in **bold** (see Table 1)

Response	A5	A6	B4	B6	B7	B8	B11	B12
wrong	**43.18**	**40.91**	31.25	**47.92**	41.67	**16.67**	31.25	31.25
correct	31.82	22.73	**25.00**	35.42	**31.25**	56.25	**39.58**	**35.42**
undetermined	25.00	36.36	43.75	16.67	27.08	27.08	29.17	33.33

Table 6

Quantitative task types: mean (a), standard deviations (b), and medians (c) of lower (l) and upper (u) bound responses in condition C ($n_3 = 47$). Except for the probability responses to task C16, all values are normalized to the value range [0, 1]. Best possible coherent bound responses are in **bold** (see Table 1)

	C1l	C1u	C2l	C2u	C3l	C3u	C4l	C4u	C5l	C5u	C6l
	0	**.70**	**.30**	**.30**	**.70**	**1**	**.70**	**1**	**.70**	**.70**	**0**
a	.42	.70	.16	.43	.25	.74	.37	.83	.63	.73	.31
b	.33	.20	.20	.36	.33	.18	.33	.22	.22	.08	.35
c	.70	.70	.00	.30	.00	.70	.30	1.00	.70	.70	.00
	C6u	C7l	C7u	C11l	C11u	C13l	C13u	C16l	C16u	C17l	C17u
	.70	**.30**	**.30**	**.30**	**.30**	**0**	**1**	**.30**	**.30**	**.30**	**.30**
a	.77	.17	.51	.14	.48	.24	.92	.31	.51	.28	.57
b	.18	.21	.40	.19	.38	.33	.21	.33	.34	.28	.34
c	.70	.00	.30	.00	.30	.00	1.00	.30	.30	.30	.70

Table 7

Qualitative task types: percentages of responses to forced tasks in condition C ($n_3 = 47$). Best possible coherent response options are in **bold** (see Table 1)

Response	C8	C9	C10	C12	C14	C15	C18	C19
wrong	**51.06**	23.40	**76.60**	**54.00**	29.79	**14.89**	17.02	12.77
true	19.15	**27.66**	8.51	20.00	**42.55**	34.04	**48.94**	**53.19**
undetermined	29.79	48.94	14.89	26.00	27.66	51.06	34.04	34.04

tests of coherence-based probability logic (e.g., [64,68,69,72]), the agreement between the predictions concerning the quantitative attack principles and the participant's responses are modest, especially in the correctness judgment conditions (A and B). For condition C, which requires to generate strengths of attacks, the majority of the participants hit some of the optimal bounds as predicted (see Table 6). In particular, looking at the median values, we observe that out of all 11 quantitative tasks of condition C, more than half of generated strengths correspond to the best possible coherent intervals in four tasks (C5, C6, C13, C16), the best possible coherent upper bounds (while most lower bounds were incoherent) in five tasks (C1, C2, C4, C7, C11), the best possible lower bound in one task (C17), and a coherent but not optimal upper bound in one task (C3). From the seven qualitative tasks in condition C, more than 50% of the responses confirmed our predictions in four tasks (i.e., C8, C10, C12, C19; see Table 7). Moreover, there is some tendency towards the predicted responses in the tasks C14 and C18.

In tasks C9 (Attacked contradiction) and C15 (Attacked tautology) around half of the participants chose "undetermined" (Table 7). These participants probably thought that it does not make sense to attack contradictions or tautologies. We observed an analogous trend in the corresponding task B4 (Attacked contradiction). In task B8 (Attacked tautology), however, most people chose "correct", as if they were just judging the truth value of a tautology (see Table 5).

Although most participants identified the coherent upper bound in the Conjunction introduction task C4, most generated an incoherent lower bound. In the corresponding correctness judgment task B2, most responses were incoherent. Thus Conjunction introduction is not corroborated by the data. Also Disjunction elimination is not supported by the data: most responses are incoherent in the corresponding tasks A2 and C3. Interestingly, most generated lower and upper bounds in the Disjunction introduction task C6 were coherent and even optimal (coincide with the best possible bounds), which supports Disjunction introduction. In the corresponding correctness judgment task B3, however, most judgments were incoherent and hence do not support Disjunction introduction.

Concerning task (**Ł.¬**), the correctness judgments in tasks A7, B1, and B5 do not support our predictions. In the corresponding strength generation tasks C2 and C11, most responses were coherent with respect to the upper bound but incoherent with respect to the lower bound. Interestingly, the task pairs B1 and B5 as well as C2 and C11 also allow to investigate the reliability of the response patterns: in both conditions the response patterns were quite similar which speaks for a good reliability. The response patterns in the (**Ł.¬**) variant tasks were also similar: (**Ł.¬**) variant is not supported in the correctness judgment task A4 and only supported with respect to the upper bound in the strength generation task C7. As in C2 and C11, most lower bound responses in C7 are incoherent.

Overall, the response tendencies in the (**C.¬**) tasks are tending in favor of supporting our predictions, with C18 being close to 50% and B11 a bit lower, close to 40%, but above chance level (1/3). This is of course at best a moderate confirmation of our predictions concerning (**C.¬**). Slightly over 50% of the responses confirm (**C.¬**) variant in task C19, which is in favor of our predictions. In the corresponding task B12, however, the response frequencies hitting the correct option is close to guessing level (1/3), which does not support (**C.¬**) variant.

The expected response options in tasks B6 and C12, which investigate Negation attack, as well as tasks A5 and C8, which investigate Negation attack′, were the most frequently chosen options. This can be seen as a moderate support of our predictions. Reflexivity was formulated identically in task A6 and task C10. However, the prediction is supported in task C10 (with 76% of the responses) whereas only 41% of the responses were as expected in task A6. This could be due to a carry-over effect of generally deeper cognitive processing in the strength generation condition C compared to the correctness judgment conditions. Concerning Contradictory attack, we observed a similar tendency: correct responses to the C14 task were more frequent (43%) compared to the B7 task 31%. The absolute support values of our predictions concerning Contradictory attack were relatively low.

The Contingent attack tasks serve to check whether people process the tasks carefully. We originally intended to test (**C.gen**), but due to a systematic error (by mistake, $C \models A$ instead of $B \models A$ was used in the tasks), we investigated what we call the *Irrelevance premise* task. This can also be seen as a consistency check. In these tasks almost all participants responded as predicted (cf. Table 1), which indicates high consistency.

Four tasks served to explore directly the connection between probability and strength of attack. Here, we observed a disparity between the predictions of ProbToAttack, AttackToProb, and AttackToProb′ and the data in conditions A and B, which required correctness judgments: the judgments did not confirm our predictions. In the corresponding tasks, which require generating strengths (i.e., C16 and C17), however, we found some support of our predictions. Specifically, task C16 which investigates AttackToProb′, the majority of participants responded as predicted. In task C17, which investigates ProbToAttack′, the majority of the lower bound responses were coherent. Again, participants scored better in the generating strengths condition compared to the correctness judgment conditions.

In sum, the results of the experiment were heterogeneous: on the one hand, some predictions of our theory were confirmed by the data; on the other hand, we observed quite some disparity between the predictions and the experimental data. In particular, we hypothesised that the tasks which involve judging the correctness of presented consequent candidates (conditions A and B) are easier and hence expected more correct responses compared to the tasks which required generating strengths of attacks (condition C). However, the data show a reversed pattern: more coherent responses were observed in condition C compared to conditions A and B. Interestingly, in the eleven quantitative strength generation tasks of condition C, most responses coincide with the optimal coherent (lower and upper) bounds in four tasks and most responses coincide with the optimal coherent upper bound in all tasks except for two tasks (i.e., task C17 where most responses are consistent with the optimal coherent lower bound and task C3, where most responses were consistent with the coherent upper bound, but not with the optimal one). Thus, the lower bounds were most frequently violated. In six out of seven qualitative tasks in condition C, we observed at least a tendency towards the predicted responses. Therefore, the data of condition C only partially confirm our predictions. Data of conditions A and B, which involved correctness judgments, did not support our predictions.

As mentioned above, the participants rated the overall clearness and the comprehensibility of the tasks on an intermediate level, which partially explains the heterogeneous results. A salient reason for not higher levels of perceived clarity of the tasks is that attack relations involve (implicit) negations. Since it is well-known in the psychology of reasoning that negations are harder to process for people compared to affirmations, we speculate that although attack relations are intuitively plausible from theoretical points of views, affirmative support relations are *psychologically* more plausible compared to attack relations. Specifically, in terms of a quantitative interpretation, modelling the support of A on F by a high conditional probability $p(F|A)$ yields an affirmative relation between A and F, while the corresponding attack relation is negative, since it requires a high conditional probability involving a negated conditioned event: $p(\neg F|A)$. Future experimental work is needed to investigate this hypothesis.

7. Attack principles and classical logical argumentation

We emphasize that our attack principles are not primarily intended for logical argumentation, where the attack relation between arguments is defined in terms of logical rather than material consequence. In fact, it is not clear at all whether a useful notion of *graded attack* should be defined as a purely logical relation. Nevertheless, at least for the qualitative case, it might be interesting to investigate for various types of attack relations that are defined as logical relations between (parts of) arguments, whether they satisfy the attack principles presented in Section 3. It is actually straightforward to specify examples of AFs that do not satisfy even straightforward attack principles, like (**C.**$\wedge$), that are clearly justified informally as well as according to coherence-based probabilistic logic. Certain constraints about the considered types of attack (defeater, undercut, etc.?), the format of arguments (e.g., consistent, minimal support?), or about the formation of arguments (e.g., can all finite sets of formulas serve as support?) may be needed in order to comply with various logical attack principles. A systematic investigation into the relation between attack principles and various forms of attacks and corresponding types of AFs is beyond the scope of this paper. However, we will at least make a few relevant observations here.

In the following we assume that an argument is of the form $\langle \Phi, A \rangle$, where Φ is a finite set of formulas such that $\Phi \models A$, i.e. Φ entails A according to classical logic. A *(logical) AF* is a pair $\langle \mathcal{A}, \longrightarrow \rangle$, where $\mathcal{A}$ is a set of arguments of the indicated form and $\longrightarrow$ is a binary relation over $\mathcal{A}$, called the *attack*

relation of the framework. Following Arieli and Strasser [6, cf. Definition 2.3, p. 75], we recall some of the better-known forms of attack relations for logical AFs.

Definition 4. Let $\alpha = \langle \Phi, A \rangle$ and $\beta = \langle \Psi, B \rangle$ be two arguments.

- α is a *defeater* of β if $A \models \neg \bigwedge_{G \in \Psi} G$.
- α is a *direct defeater* of β if there is a $G \in \Psi$ such that $\Phi \models \neg G$.
- α is an *undercut* of β if there is $\Psi' \subseteq \Psi$ such that $A \models \neg \bigwedge_{F \in \Psi'} F$ and $\neg \bigwedge_{F \in \Psi'} F \models A$.
- α is a *direct undercut* of β if there is a $G \in \Psi$ such that $A \models \neg G$ and $\neg G \models A$.
- α is a *canonical undercut* of β if $A \models \neg \bigwedge_{F \in \Psi} F$ and $\neg \bigwedge_{F \in \Psi} F \models A$.
- α is a *rebuttal* of β if $A \models \neg B$ and $\neg B \models A$.
- α is a *defeating rebuttal* of β if $A \models \neg B$.

Definition 5. Let $\Lambda = \langle \mathcal{A}, \longrightarrow \rangle$ be a logical AF. We say that Λ is *based on (direct) defeat / (direct/canonical) undercut / (defeating) rebuttal* if for all arguments $\alpha, \beta \in \mathcal{A}$: $\alpha \longrightarrow \beta$ iff α is a (direct) defeater / (direct/canonical) undercut / (defeating) rebuttal of β, respectively.

In investigating which attack principles of Section 3 are satisfied for which types of attacks, we start with (defeating) rebuttal, which ignores the support part of arguments. Obviously, (unqualified) rebuttal between arguments is a symmetric relation. Therefore, rebuttal cannot reflect the fact that an argument against, e.g., $A \wedge B$ implicitly entails the existence an argument against A as well as one against B. In other words, principles like (**C.**$\wedge$) trivially fail for AFs that are (solely) based on rebuttal. If, however, we consider the more general attack relation of defeating rebuttal, we obtain the following.

Proposition 10. *If an AF Λ is based on defeating rebuttal, then the attack principles* (**C.**$\wedge$)*,* (**C.**$\vee$)*,* (**C.**$\vee$)**'***,* (**C.** $\supset$)*,* (**C.**$\neg$)*, and* (**C.**$\bot$) *are satisfied.*[13]

Proof. Let $\Lambda = \langle \mathcal{A}, \longrightarrow \rangle$. Assume that there is an argument φ with claim F in $\mathcal{A}$ that is a defeating rebuttal of some argument α with claim A or of some argument β with claim B, where $\alpha, \beta \in \mathcal{A}$. This implies that $F \models \neg A$ or $F \models \neg B$. In both cases it follows that $F \models \neg(A \wedge B)$. This means that φ is also a defeating rebuttal of any argument featuring claim $A \wedge B$. Hence (**C.**$\wedge$) is satisfied.

The proofs for (**C.**$\vee$), (**C.**$\vee$)**'**, and (**C.** $\supset$) are similar to that for (**C.**$\wedge$). (**C.** $\supset$) is satisfied since $F \models \neg(A \supset B)$ entails $F \models \neg B$. For (**C.**$\vee$) it suffices to observe that $F \models \neg(A \vee B)$ entails both $F \models \neg A$ and $F \models \neg B$. But also the inverse entailment holds. Hence also (**C.**$\vee$)**'** is satisfied.

The case for (**C.**$\neg$) is somewhat different. (**C.**$\neg$), applied to defeating rebuttal, asserts that $F \models \neg A$ does not entail $F \not\models A$. This is indeed the case, since F is required to be non-contradictory.

Finally, note every argument is a defeating rebuttal of any argument featuring a contradictory claim, since $F \models \neg\bot$, for every F. This means that (**C.**$\bot$) is satisfied as well. □

Proposition 11. *In AFs based on either rebuttal or defeating rebuttal the attack principles* (**C.**$\wedge$)**'***,* (**C.** $\supset$)**'***, and* (**C.**$\neg$)**'** *are not satisfied in general.*

Proof. For a counterexample to (**C.**$\wedge$)**'**, consider an AF where all claims of arguments are either p, q, $p \wedge q$, or $\neg(p \wedge q)$, for two distinct propositional variables p and q. Clearly the arguments with claims

[13] Recall from Section 3 that the attack principles refer only to AFs, in which formulas that have the logical form indicated in the attack principles actually occur as claims of arguments.

$p \wedge q$ and $\neg(p \wedge q)$, respectively, rebut each other. Hence, in particular, there is an argument attacking $p \wedge q$. But there is no argument that is a (defeating) rebuttal of an argument with claim p or with claim q. This means that (**C.**$\wedge$)**'** is not satisfied.

A similar counterexample to (**C.** $\supset$)**'** is obtained by considering an AF where the only claims of arguments are p, $\neg p$, and $p \supset q$, respectively.

That (**C.**$\neg$)**'** is not satisfied follows from the observation that $F \models \neg\neg A$ does not follow from $F \not\models \neg A$ in general. □

Propositions 10 and 11 are largely in agreement with the classification of attack principles according to our probabilistic semantics in Section 4. The only possible exception concerns principle (**C.**$\vee$)**'**. Note however that, as pointed out in Proposition 1, (**C.**$\vee$)**'** also holds with respect to the probabilistic interpretation for the particular case where the threshold t is set to $t = 1$. Hence, modulo that specific case, defeating rebuttal for logical AFs complies with our attack principles. We also noted that for unqualified rebuttal already (**C.**$\wedge$) fails. This means that one should generalize rebuttal to defeating rebuttal if one wants to respect the straightforward existence of further attacks, like that on $A \wedge B$ whenever there is one for A.

Let us now turn to defeat and undercut. Any non-trivial principle about the existence of further attacking arguments in presence of certain defeats or undercuts will have to make some assumptions about the formation of the set of arguments in a given AF.

Definition 6. An argument $\beta = \langle \Psi, B \rangle$ *arises from augmenting the support* of argument $\alpha = \langle \Phi, A \rangle$ if $\Phi \subseteq \Psi$.

Definition 7. We say that an AF $\Lambda = \langle \mathcal{A}, \longrightarrow \rangle$ satisfies *support augmentation* if the following condition holds. If $\mathcal{A}$ contains arguments claiming A and B, respectively, where $B \models A$, then at least one of the arguments with claim B arises from augmenting the support of some argument for A in $\mathcal{A}$.

Proposition 12. *If an argumentation framework* Λ *satisfies support augmentation and is based on defeat, direct defeat, undercut, direct undercut or canonical undercut, then the attack principles* (**C.**$\wedge$)*,* (**C.**$\vee$)*, and* (**C.** $\supset$) *are satisfied.*

Proof. Let $\Lambda = \langle \mathcal{A}, \longrightarrow \rangle$. Suppose that there is an argument φ with claim F in Λ that defeats some argument $\langle \Psi, A \rangle$. This implies that $F \models \neg \bigwedge_{G \in \Psi} G$. Since Λ satisfies support augmentation and since $A \wedge B \models A$, $\mathcal{A}$ must also contain an argument γ of the form $\langle \Psi', A \wedge B \rangle$, where $\Psi \subseteq \Psi'$. But from $\Psi \subseteq \Psi'$ and $F \models \neg \bigwedge_{G \in \Psi} G$ it follows that $F \models \neg \bigwedge_{G \in \Psi'} G$. This means that φ is also a defeater of γ. (The case where φ defeats an argument for B, instead of one for A, is analogous.) Hence (**C.**$\wedge$) is satisfied for defeat.

The above argument for defeaters straightforwardly generalizes to AFs based on direct defeat, undercut, direct undercut or canonical undercut. It suffices to observe that the argument remains valid if we refer to either a subset or an element of the support set Φ, rather than to Φ itself, and that it also does not matter whether we assume that F entails the corresponding negated formula or is logically equivalent to it.

The proofs for (**C.**$\vee$) and (**C.** $\supset$) are very similar to that for (**C.**$\wedge$): support augmentation, jointly with $A \models A \vee B$, $B \models A \vee B$, $B \models A \supset B$, allows one to establish the existence of the required attacking arguments in each case. □

The principle (**C.**¬) is not satisfied for defeat and undercut without imposing further conditions. Regarding (**C.**⊥), it is straightforward to see that every argument is a defeater and as well as an undercut of every argument featuring a contradictory claim. But is also clear that (**C.**⊥) may be violated if only *direct* defeat and *direct* undercut is considered. Finally, we remark in passing that one can construct AFs that satisfy support augmentation that also satisfy some of those attack principles that should be discarded according to our probabilistic interpretation. However, as mentioned above, a systematic investigation of the relation between logical attack principles and types of logical AFs is beyond the scope of this paper. Such an investigation is rather a topic for further research that will, e.g., have to include discussions about various forms of argument formation, including the role of minimality and consistency constraints on the support part of logical arguments.

8. Relations to computational argumentation and AI

There are many contributions in the vast growing field of argumentation and AI [8,11,75] that mention probabilities and weighted attacks in various different ways. However, those approaches, following Dung's paradigm for computational argumentation, focus on aspects of argumentation that are at best indirectly related to our current concerns. Given the prominence of Dung style argumentation theory, it may still be beneficial to briefly discuss some of this work and to highlight potential relations to our investigation of logical attack principles.

First, we point out that we follow Dung's approach to argumentation in focusing on attack, rather than on support between arguments. While, more recently, various authors (see, e.g., [13,19]) have suggested to add also an explicit support relation to the AFs, we decided to pay respect to the observation that models of argumentation should give prominence to the interaction between arguments and counter-arguments and thus to the attack relation between arguments. Nevertheless, we emphasize that the probabilistic semantics of the attack relation presented in this paper can straightforwardly be adapted to an interpretation of support between arguments. Corresponding 'support principles', arising from the logical form of claims of arguments in analogy to our logical attack principles, can readily be formulated. With an eye to the experimental part, we suppose that many such support principles are actually easier to judge as either intuitively valid or invalid than the attack principles investigated in Section 6, since they do not involve negating attacked claims. This clearly is a subject matter for future research.

It is customary to distinguish explicitly between abstract argumentation and logic-based argumentation in the computational argumentation community. (This can be traced back to Dung [30]; [2] and [47] are just two of many more recent papers, where the distinction is made explicit already at the outset.) The focus on the logical form of argumentative claims seems to place our investigation firmly in the field of logic-based argumentation. However, as already pointed out in Section 2, our decision to look only at the logical form of claims and disregard the formal structure of the support part of arguments entirely, places the corresponding principles on a level that is intermediary between abstract and logical argumentation, i.e., what we called the semi-abstract level. This implies that our results do not depend on a particular version of logic-based argumentation. Thus, our probabilistic interpretation of the attack relation can, in principle, be applied to quite different formats of logical arguments, like, e.g., the one suggested in [11], the more complex format used in ASPIC$^+$ [61], or sequent-based formats [6,83]. This has the advantage that we do not have to engage into the ongoing debate about the appropriateness of certain restrictions on the support of arguments, like minimality or consistency (see, e.g., [6,25]).

The investigated logical attack principles can be viewed as rationality postulates. However, the later expression is often used in a somewhat different sense in the literature on Dung-style AFs. For example, Amgoud in [2] proposes five rationality postulates that logic-based argumentation system should satisfy. Similar postulates have been proposed in [14] and [15]. Note that those postulates do not refer to the logical form of arguments, but rather call for *global* properties of the framework, like, e.g., that the set of all considered arguments should be closed under sub-arguments or that the claims of arguments that are members of Dung-style 'extensions' should be jointly consistent. In contrast, our principles are *local*, in the sense that they postulate the attack (or lack of attack) between arguments featuring claims of a certain logical form. Gorogiannis and Hunter [47] formulate 'postulates concerning attack functions' that, while not pertaining to the logical form of claims, can be classified as local, as well. In particular, the principle called (D2') in [47] is very close to the general attack principle (**C.gen**), mentioned in Section 3. The only difference is that, in our terminology, F in $F \longrightarrow A$ does not denote a particular argument, but refers to any argument featuring the claim F. However, all the above mentioned rationality principles are qualitative, not quantitative. Moreover, there is a clear divergence of motivation and interest between our contribution and that of researchers working in the paradigm originating with Dung [30]. While most of the latter research community typically focuses on effective computational extraction of information from arguments automatically compiled from large, inconsistent data bases, we are interested in the interpretation and justification of the attack relation between arguments, as they appear in human discourse. Nevertheless we hope that our study may have repercussions also for computational argumentation, since it is important for computer-based reasoning systems to pay attention to the human interpretability of underlying reasoning principles. Obviously, this is a particularly challenging task for quantitative principles, like those investigated in this paper.

As already mentioned, various forms of quantitative AFs have been suggested in the literature, see, e.g., [1,3–5,9,24,32,60]. There is no consensus on whether one should directly put weights on individual arguments or, rather consider degrees of strength of attacks between arguments in the first place (see [32] for a discussion of this issue). We follow the second approach here and address questions that are so far neglected in the literature: How should the weights (degrees of strength) of attacks be interpreted systematically? How do they interact with the logical form of argumentative claims? Which corresponding principles are readily accepted by human reasoners? In this manner we hope to contribute at least indirectly to the fast growing literature on weighted AFs.

There is also a considerable amount of literature on probability-based approaches to argumentation in AI, see, e.g., [28,31,53,55,56,59,76]. Following Hunter [53], two main approaches in this area are (1) the *constellations approach*, modelling uncertainty in the topology of the argument framework by considering probability distributions over possible argument graphs, and (2) the *epistemic approach*, where one attaches degrees of belief to arguments. Related to the second approach, the connection between support and claim of arguments is sometimes endowed with uncertainty measures, like conditional probabilities (see, e.g., [26]), possibility measures [17], or coarser grades of uncertainty (see, e.g., [38,39]). Somewhat closer to our concern is an extension of the epistemic approach, presented in [73], where degrees of beliefs are associated with individual attacks. However, all mentioned approaches aim at a different target than ours: they consider generalizations of Dung's AFs by associating probabilities either to arguments or to attacks between arguments, resulting in probability distributions either over the arguments or over subgraphs of AFs, respectively. Similarly to the literature on weighted AFs mentioned above, the focus is on global effects on the acceptability of sets of arguments in the framework, whereas our use of probability operates on a different level, serving as semantic tool to better understand the plausibility of (local) constraints on the attack relation induced by the logical form of attacked arguments.

9. Concluding remarks

We showed how the coherence approach to probability can serve to guide the rational selection of qualitative and quantitative principles regarding the existence of attacks on logically compound claims. More research is needed to deepen and to generalize our formal results: e.g., by interpreting implication as conditional probability (or as previsions in conditional random quantities) or by generalizations to fuzzy events. We also presented an experiment to explore the psychological plausibility of selected features of our approach. While we are convinced that our approach is intuitive and plausible from a theoretical point of view, we were surprised by the relatively heterogeneous experimental results. We observed some evidence in favor of our hypotheses under the experimental condition where participants generated strengths of attacks. Interestingly, the majority of the participants hit some of the optimal coherent bounds as predicted. Violations most frequently concerned the lower bounds. When the participants merely judged the correctness of attack strength candidates, however, most responses did not confirm our hypotheses. The heterogeneous agreement between the predictions and the responses could be caused by various factors including (i) lower data quality in a lecture hall experiment compared to individual testing, (ii) different response formats (the open response format (strength generation) appeared to be more appropriate compared to the forced choice response format (correctness judgments) to investigate quantitative attack principles), and (iii) possible confusions caused by the negations involved in the probabilistic semantics of the attack relations (i.e., $p(\neg B|A)$ should be high in order that $A \longrightarrow B$ holds). Although attack relations are intuitive and plausible from theoretical points of views, maybe support relations are *psychologically* more intuitive, as they can be represented positively by the human mind without requiring implicit negations. Future experimental work is needed to further explore the psychological plausibility of formal attack principles.

Moreover, it has been pointed out (see, e.g., [74]) that epistemic contexts may trigger sceptical reasoning while practical contexts trigger credulous reasoning. To what extent peoples' judgments of attack strength are context-dependent in this sense is another topic for future experimental investigations.

Our attempt of giving a positive description of rationality postulates for quantitative attack principles is related to probabilistic semantics of nonmonotonic reasoning: on the one hand, for example, premise strengthening, contraposition, and transitivity neither hold in nonmonotonic reasoning nor in our framework. On the other hand, the nonmonotonic reasoning rules like those of System P [41,58] or Weak Transitivity [42] also hold in our framework. Moreover, the attack relation formalized as $p(\neg B|A) \geqslant x$ can also be interpreted as a formalization of the attack of the normality condition of a corresponding rule: *by default, if A, then B*.

Researchers working in the Dung style tradition might be interested in the question how the various types of extensions (grounded, preferred, stable, etc.) for AFs are affected by imposing or rejecting logical attack principles like the ones investigated here. While this is certainly an interesting question from the point of view of computational argumentation, it is quite removed from our focus on the interpretation and justification of logical attack principles. We nevertheless hope that it will be tackled in future research.

In Section 7 we made some observations about the special case of *logical* AFs, where attack is defined in terms of logical entailment. It turned out that qualitative attack principles that are justified according to our probabilistic semantics are automatically satisfied for defeating rebuttal as attack relation. Moreover, principles that are not probabilistically justified are, in general, not satisfied by AFs based on defeating rebuttal. However, other logical attack relations do not readily fall in line with our classification of probabilistically plausible and implausible attack principles. This calls for further investigations

on appropriate attack principles for logical argumentation. In particular, it remains to be investigated whether and how it can be justified that many forms of logical attack do not comply with intuitively plausible principles like (**C.**$\wedge$).

In our paper we used classical logic for the qualitative attack principles. In particular regarding implication and negation it would be interesting to use a different logic, like relevance logic or a nonmonotonic logic, to form arguments. This in turn triggers further questions about corresponding attack principles and their interpretation.

As mentioned above, it is natural to look at rationality principles for support relations between arguments in addition to looking at attack relations. For example, consider that arguments with claim F support A as well as B, then it seems natural to infer that F also supports $A \wedge B$. Moreover, some principles may combine support and attack relations, e.g.: if F supports A but attacks B, then F attacks $A \supset B$. We will investigate such principles from qualitative, quantitative, and experimental points of views in future work.

Acknowledgements

Thanks to three anonymous referees, Barbara Vantaggi, and Gernot Kleiter for useful comments. We also thank Gernot Salzer for making the experiment possible during his class and the students who participated.

Niki Pfeifer was supported by his BMBF project 01UL1906X.

References

[1] T. Alsinet, R. Béjar, L. Godo and F. Guitart, RP-DeLP: A weighted defeasible argumentation framework based on a recursive semantics, *Journal of Logic and Computation* **26**(4) (2016), 1315–1360. doi:10.1093/logcom/exu008.

[2] L. Amgoud, Postulates for logic-based argumentation systems, *International Journal of Approximate Reasoning* **55**(9) (2014), 2028–2048. doi:10.1016/j.ijar.2013.10.004.

[3] L. Amgoud and J. Ben-Naim, Weighted bipolar argumentation graphs: Axioms and semantics, in: *Twenty-Seventh International Joint Conference on Artificial Intelligence – IJCAI 2018*, 2018.

[4] L. Amgoud, J. Ben-Naim, D. Doder and S. Vesic, Acceptability semantics for weighted argumentation frameworks, in: *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017, pp. 56–62.

[5] L. Amgoud and D. Doder, Gradual semantics for weighted graphs: An unifying approach, in: *Sixteenth International Conference on Principles of Knowledge Representation and Reasoning*, 2018.

[6] O. Arieli and C. Straßer, Sequent-based logical argumentation, *Argument & Computation* **6**(1) (2015), 73–99. doi:10.1080/19462166.2014.1002536.

[7] O. Arieli and C. Straßer, On minimality and consistency tolerance in logical argumentation frameworks, in: *Computational Models of Argument: Proceedings of COMMMA 2020*, H. Prakken, S. Bistarelli, F. Santini and C. Taticchi, eds, IOS Press, 2020, pp. 91–102.

[8] P. Baroni, D.M. Gabbay, M. Giacomin and L. van der Torre, *Handbook of Formal Argumentation*, College Publications, 2018.

[9] P. Baroni, A. Rago and F. Toni, How many properties do we need for gradual argumentation? in: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18)*, New Orleans, Louisiana, USA, February 2–7, 2018, S.A. McIlraith and K.Q. Weinberger, eds, AAAI Press, 2018, pp. 1736–1743.

[10] T.J.M. Bench-Capon and P.E. Dunne, Argumentation in artificial intelligence, *Artificial Intelligence* **171** (2007), 619–641. doi:10.1016/j.artint.2007.05.001.

[11] P. Besnard and A. Hunter, *Elements of Argumentation*, MIT Press, Cambridge, 2008.

[12] V. Biazzo and A. Gilio, A generalization of the fundamental theorem of de Finetti for imprecise conditional probability assessments, *International Journal of Approximate Reasoning* **24**(2–3) (2000), 251–272. doi:10.1016/S0888-613X(00)00038-4.

[13] G. Boella, D.M. Gabbay, L. van der Torre and S. Villata, Support in abstract argumentation, in: *Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10)*, Frontiers in Artificial Intelligence and Applications, IOS Press, 2010, pp. 40–51.
[14] M. Caminada, Rationality postulates: Applying argumentation theory for non-monotonic reasoning, *Journal of Applied Logics* **4**(8) (2017), 2707–2734.
[15] M. Caminada and L. Amgoud, On the evaluation of argumentation formalisms, *Artificial Intelligence* **171**(5–6) (2007), 286–310. doi:10.1016/j.artint.2007.02.003.
[16] F. Cerutti, M. Cramer, M. Guillaume, E. Hadoux, A. Hunter and S. Polberg, Empirical cognitive studies about formal argumentation, in: *Handbook of Formal Argumentation (Volume 2)*, D. Gabbay, M. Giacomin, G.R. Simari and M. Thimm, eds, College Publications, in press.
[17] C.I. Chesñevar, G.R. Simari, L. Godo and T. Alsinet, Argument-based expansion operators in possibilistic defeasible logic programming: Characterization and logical properties, in: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference, ECSQARU 2005*, Barcelona, Spain, July 6–8, 2005, Proceedings, L. Godo, ed., Lecture Notes in Computer Science, Vol. 3571, Springer, 2005, pp. 353–365.
[18] P. Cintula, C.G. Fermüller and C. Noguera, Fuzzy logic, in: *The Stanford Encyclopedia of Philosophy*, E.N. Zalta, ed., Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/win2021/entries/logic-fuzzy/.
[19] A. Cohen, S. Gottifredi, A.J. García and G.R. Simari, A survey of different approaches to support in argumentation systems, *The Knowledge Engineering Review* **29**(5) (2014), 513. doi:10.1017/S0269888913000325.
[20] G. Coletti and R. Scozzafava, *Probabilistic Logic in a Coherent Setting*, Kluwer, 2002.
[21] E.A. Corsi, Argumentation theory and alternative semantics for non-classical logics, PhD thesis, TU Wien, 2021.
[22] E.A. Corsi and C.G. Fermüller, Logical argumentation principles, sequents, and nondeterministic matrices, in: *Logic, Rationality, and Interaction: 6th International Workshop, LORI 2017*, Sapporo, Japan, September 11–14, 2017, Proceedings, A. Baltag, J. Seligman and T. Yamada, eds, LNCS, Vol. 10455, Springer, Berlin, 2017, pp. 422–437.
[23] E.A. Corsi and C.G. Fermüller, Connecting fuzzy logic and argumentation frames via logical attack principles, *Soft Computing* **23** (2019), 2255–2270. doi:10.1007/s00500-018-3513-2.
[24] S. Coste-Marquis, S. Konieczny, P. Marquis and M.A. Ouali, Weighted attacks in argumentation frameworks, in: *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*, AAAI Press, 2012, pp. 593–597.
[25] M. D'Agostino and S. Modgil, Classical logic, argument and dialectic, *Artificial Intelligence* **262** (2018), 15–51. doi:10.1016/j.artint.2018.05.003.
[26] P. Dellunde, L. Godo and A. Vidal, Probabilistic argumentation: An approach based on conditional probability – A preliminary report, in: *Logics in Artificial Intelligence – 17th European Conference, JELIA 2021*, May 17–20, 2021, Virtual Event, Proceedings, W. Faber, G. Friedrich, M. Gebser and M. Morak, eds, Lecture Notes in Computer Science, Vol. 12678, Springer, 2021, pp. 25–32.
[27] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics* **38** (1967), 325–339. doi:10.1214/aoms/1177698950.
[28] D. Doder and S. Woltran, Probabilistic argumentation frameworks – A logical approach, in: *International Conference on Scalable Uncertainty Management*, Springer, 2014, pp. 134–147. doi:10.1007/978-3-319-11508-5_12.
[29] D. Dubois and H. Prade, *Possibility Theory. An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.
[30] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intelligence* **77**(9) (1995), 321–357. doi:10.1016/0004-3702(94)00041-X.
[31] P.M. Dung and P.M. Thang, Towards (probabilistic) argumentation for jury-based dispute resolution, *COMMA* **216** (2010), 171–182.
[32] P.E. Dunne, A. Hunter, P. McBurney, S. Parsons and M. Wooldridge, Weighted argument systems: Basic definitions, algorithms, and complexity results, *Artificial Intelligence* **175**(2) (2011), 457–486. doi:10.1016/j.artint.2010.09.005.
[33] P.E. Dunne and M. Wooldridge, Complexity of abstract argumentation, in: *Argumentation in Artificial Intelligence*, G.R. Simari and I. Rahwan, eds, Springer, 2009, pp. 85–104. doi:10.1007/978-0-387-98197-0_5.
[34] C. Dutilh Novaes, Argument and argumentation, in: *The Stanford Encyclopedia of Philosophy*, E.N. Zalta, ed., Metaphysics Research Lab, Stanford University, 2021.
[35] W. Dvořák and S. Woltran, Complexity of abstract argumentation under a claim-centric view, *Artificial Intelligence* **285** (2020), 103290. doi:10.1016/j.artint.2020.103290.
[36] J.S.B.T. Evans, *The Psychology of Deductive Reasoning*, Routledge, London, 1982.
[37] J.S.B.T. Evans, J.L. Allen, S. Newstead and P. Pollard, Debiasing by instruction: The case of belief bias, *European Journal of Cognitive Psychology* **6** (1994), 263–285. doi:10.1080/09541449408520148.
[38] J. Fox, Arguing about the evidence: A logical approach, in: *Proceedings of the British Academy*, Vol. 171, 2011, p. 151.

[39] J. Fox and S. Parsons, Arguing about beliefs and actions, in: *Applications of Uncertainty Formalisms*, A. Hunter and S. Parsons, eds, Lecture Notes in Computer Science, Vol. 1455, Springer, 1998, pp. 266–302. doi:10.1007/3-540-49426-X_13.
[40] G. Gentzen, Untersuchungen über das logische Schließen, *Mathematische Zeitschrift* **39** (1935), 176–210, 405–431. doi:10.1007/BF01201363.
[41] A. Gilio, Probabilistic reasoning under coherence in System P, *Annals of Mathematics and Artificial Intelligence* **34** (2002), 5–34. doi:10.1023/A:1014422615720.
[42] A. Gilio, N. Pfeifer and G. Sanfilippo, Transitivity in coherence-based probability logic, *Journal of Applied Logic* **14** (2016), 46–64. doi:10.1016/j.jal.2015.09.012.
[43] A. Gilio, N. Pfeifer and G. Sanfilippo, Probabilistic entailment and iterated conditionals, in: *Logic and Uncertainty in the Human Mind: A Tribute to David E. Over*, S. Elqayam, I. Douven, J.S.B.T. Evans and N. Cruz, eds, Routledge, London, 2020, pp. 71–101.
[44] A. Gilio and G. Sanfilippo, Conditional random quantities and compounds of conditionals, *Studia Logica* **102**(4) (2014), 709–729. doi:10.1007/s11225-013-9511-6.
[45] A. Gilio and G. Sanfilippo, Generalized logical operations among conditional events, *Applied Intelligence* **49**(1) (2019), 79–102. doi:10.1007/s10489-018-1229-8.
[46] A. Gilio and G. Sanfilippo, Compound conditionals, Fréchet–Hoeffding bounds, and Frank t-norms, *International Journal of Approximate Reasoning* **136** (2021), 168–200. doi:10.1016/j.ijar.2021.06.006.
[47] N. Gorogiannis and A. Hunter, Instantiating abstract argumentation with classical logic arguments: Postulates and properties, *Artificial Intelligence* **175**(9–10) (2011), 1479–1497. doi:10.1016/j.artint.2010.12.003.
[48] D. Grooters and H. Prakken, Two aspects of relevance in structured argumentation: Minimality and paraconsistency, *Journal of Artificial Intelligence Research* **56** (2016), 197–245. doi:10.1613/jair.5058.
[49] R. Haenni, Probabilistic argumentation, *Journal of Applied Logic* **7** (2009), 155–176. doi:10.1016/j.jal.2007.11.006.
[50] U. Hahn and M. Oaksford, The rationality of informal argumentation: A Bayesian approach to reasoning fallacies, *Psychological Review* **114**(3) (2007), 704–732. doi:10.1037/0033-295X.114.3.704.
[51] P. Hájek, *Metamathematics of Fuzzy Logic*, Kluwer, Dordrecht, 1998.
[52] C.L. Hamblin, *Fallacies*, Methuen, London, 1970.
[53] A. Hunter, A probabilistic approach to modelling uncertain logical arguments, *International Journal of Approximate Reasoning* **54**(1) (2013), 47–81. doi:10.1016/j.ijar.2012.08.003.
[54] A. Hunter, Argument strength in probabilistic argumentation based on defeasible rules, *International Journal of Approximate Reasoning* **146** (2022), 79–105. doi:10.1016/j.ijar.2022.04.003.
[55] A. Hunter and M. Thimm, Probabilistic argumentation with incomplete information, in: *ECAI*, 2014, pp. 1033–1034.
[56] A. Hunter and M. Thimm, Probabilistic reasoning with abstract argumentation frameworks, *Journal of Artificial Intelligence Research* **59** (2017), 565–611. doi:10.1613/jair.5393.
[57] G.D. Kleiter, A.J.B. Fugard and N. Pfeifer, A process model of the understanding of uncertain conditionals, *Thinking & Reasoning* **24**(3) (2018), 386–422. doi:10.1080/13546783.2017.1422542.
[58] S. Kraus, D. Lehmann and M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, *Artificial Intelligence* **44** (1990), 167–207. doi:10.1016/0004-3702(90)90101-5.
[59] H. Li, N. Oren and T.J. Norman, Probabilistic argumentation frameworks, in: *International Workshop on Theory and Applications of Formal Argumentation*, Springer, 2011, pp. 1–16.
[60] D.C. Martınez, A.J. Garcıa and G.R. Simari, An abstract argumentation framework with varied-strength attacks, in: *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, 2008, pp. 135–144.
[61] S. Modgil and H. Prakken, The ASPIC+ framework for structured argumentation: A tutorial, *Argument & Computation* **5**(1) (2014), 31–62. doi:10.1080/19462166.2013.869766.
[62] M. Oaksford, N. Chater and U. Hahn, Human reasoning and argumentation: The probabilistic approach, in: *Reasoning: Studies of Human Inference and Its Foundations*, J. Adler and L. Rips, eds, Cambridge University Press, Cambridge, 2008.
[63] S. Parsons, Normative argumentation and qualitative probability, in: *Qualitative and Quantitative Practical Reasoning*, D.M. Gabbay, R. Kruse, A. Nonnengart and H.J. Ohlbach, eds, Springer, Berlin, 1997, pp. 466–480. doi:10.1007/BFb0035642.
[64] N. Pfeifer, The new psychology of reasoning: A mental probability logical perspective, *Thinking & Reasoning* **19**(3–4) (2013), 329–345. doi:10.1080/13546783.2013.838189.
[65] N. Pfeifer, On argument strength, in: *Bayesian Argumentation. The Practical Side of Probability*, F. Zenker, ed., Synthese Library (Springer), Dordrecht, 2013, pp. 185–193. doi:10.1007/978-94-007-5357-0_10.
[66] N. Pfeifer, Reasoning about uncertain conditionals, *Studia Logica* **102**(4) (2014), 849–866. doi:10.1007/s11225-013-9505-4.

[67] N. Pfeifer, Probability logic, in: *Handbook of Rationality*, M. Knauff and W. Spohn, eds, The MIT Press, Cambridge, MA, in press.
[68] N. Pfeifer and G.D. Kleiter, Coherence and nonmonotonicity in human reasoning, *Synthese* **146**(1–2) (2005), 93–109. doi:10.1007/s11229-005-9073-x.
[69] N. Pfeifer and G.D. Kleiter, Framing human inference by coherence based probability logic, *Journal of Applied Logic* **7**(2) (2009), 206–217. doi:10.1016/j.jal.2007.11.005.
[70] N. Pfeifer and H. Pankka, Modeling the Ellsberg paradox by argument strength, in: *Proceedings of the 39th Cognitive Science Society Meeting*, Austin, TX, G. Gunzelmann, A. Howes, T. Tenbrink and E. Davelaar, eds, The Cognitive Science Society, 2017, pp. 2888–2893.
[71] N. Pfeifer and G. Sanfilippo, Probabilistic squares and hexagons of opposition under coherence, *International Journal of Approximate Reasoning* **88** (2017), 282–294. doi:10.1016/j.ijar.2017.05.014.
[72] N. Pfeifer and L. Tulkki, Conditionals, counterfactuals, and rational reasoning. An experimental study on basic principles, *Minds and Machines* **27**(1) (2017), 119–165. doi:10.1007/s11023-017-9425-6.
[73] S. Polberg, A. Hunter and M. Thimm, Belief in attacks in epistemic probabilistic argumentation, in: *International Conference on Scalable Uncertainty Management*, Springer, 2017, pp. 223–236. doi:10.1007/978-3-319-67582-4_16.
[74] H. Prakken, Combining sceptical epistemic reasoning with credulous practical reasoning, in: *Computational Models of Argument*, P.E. Dunne and T.J.M. Bench-Capon, eds, IOS Press, Amsterdam, 2006, pp. 311–322.
[75] I. Rahwan and G.R. Simari, *Argumentation in Artificial Intelligence*, Vol. 47, Springer, 2009.
[76] R. Riveret, P. Baroni, Y. Gao, G. Governatori, A. Rotolo and G. Sartor, A labelling framework for probabilistic argumentation, *Annals of Mathematics and Artificial Intelligence* **83**(1) (2018), 21–71. doi:10.1007/s10472-018-9574-1.
[77] G. Sanfilippo, A. Gilio, D.E. Over and N. Pfeifer, Probabilities of conditionals and previsions of iterated conditionals, *International Journal of Approximate Reasoning* **121** (2020), 150–173. doi:10.1016/j.ijar.2020.03.001.
[78] G. Sanfilippo, N. Pfeifer and A. Gilio, Generalized probabilistic modus ponens, in: *ECSQUARU 2017*, A. Antonucci, L. Cholvy and O. Papini, eds, LNCS, Vol. 10369, Springer, 2017, pp. 480–490.
[79] G. Sanfilippo, N. Pfeifer, D.E. Over and A. Gilio, Probabilistic inferences from conjoined to iterated conditionals, *International Journal of Approximate Reasoning* **93** (2018), 103–118. doi:10.1016/j.ijar.2017.10.027.
[80] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
[81] W. Spohn, Ordinal conditional functions: A dynamic theory of epistemic states, in: *Causation in Decision, Belief Change, and Statistics*, W. Harper and B. Skyrms, eds, Reidel, Dordrecht, 1988, pp. 105–134. doi:10.1007/978-94-009-2865-7_6.
[82] K. Stenning and M. van Lambalgen, *Human Reasoning and Cognitive Science*, The MIT Press, Cambridge, MA, 2008.
[83] C. Straßer and O. Arieli, Normative reasoning by sequent-based argumentation, *Journal of Logic and Computation* **29**(3) (2019), 387–415. doi:10.1093/logcom/exv050.
[84] S.E. Toulmin (ed.), *The Uses of Argument*, Cambridge University Press, Cambridge, 2003.
[85] F.H. van Eemeren, B. Grassen, E.C.W. Krabbe, F. Snoeck Henkemans, B. Verheij and J.H.M. Wagemans, *Handbook of Argumentation Theory*, Springer, Dordrecht, 2014.
[86] D. Walton, C. Reed and F. Macagno, *Argumentation Schemes*, Cambridge University Press, 2008.
[87] F. Zenker (ed.), *Bayesian Argumentation: The Practical Side of Probability*, Synthese Library (Springer), Dordrecht, 2013.