Generalizing Probabilistic Material Implication and Bayesian Conditionals

Michael Jahn and Matthias Scheutz

[a]Computer Science, Tufts University, 177 College Ave, Medford, 02155, MA, USA

Abstract

Conditional statements in natural language of the form "if A then B" have multiple interpretations that require different logical treatment. In this paper, we focus on probabilistic "if A then B" rules that are given either a Bayesian interpretation via conditional probabilities $\boldsymbol{P}(B \mid A)$, or couched as probabilistic material implication. While some have argued that Bayesian conditionals are the correct way to think about such rules, there are challenges with standard inferences such as *modus ponens* and *modus tollens* that might make probabilistic material implication a better candidate at times for rule-based systems employing forward-chaining; and arguably material implication is still suitable when information about prior or conditional probabilities is not available at all. We investigate a generalization of probabilistic material implication and Bayesian conditionals that combines the advantages of both formalisms in a systematic way and prove basic properties of the generalized implication, in particular, various bounds as well as properties of inference chains in graphs. And most importantly we provide a novel and natural interpretation of the generalized implication's main parameter.

Keywords:

1. Introduction

Formal rule-based inference is widely used in a variety of fields, from artificial intelligence, to philosophy. Yet, "if-then" rules, or *conditional statements*, cover a wide range of logically distinct concepts and relations such as *analytical truths* ("if p is true, then p $\vee$ q is true"), *concept relationships* ("if x is an apple, then x is a fruit"), *causal relations* ("if the power is off, the light will not turn on"), *abductive inferences* ("if the light does not turn on, then the power must be off"), *normative constraints* ("if the sign says 'no smoking', smoking is not allowed"), *hypotheticals* ("if the plane has a mechanical problem, the flight will be cancelled"), and *counterfactuals* ("if the mechanical issue had been fixed, the plane could have departed") (see also [7, 37, 9, 29, 14]).

Given the conceptual diversity covered by natural language "if-then" expressions, it is not not suprising that there is still disagreement on how to best capture the *probabilistic rules* that extend them. These have traditionally been modeled either with material implication (the "logical" approach), or with a conditional probability (the "Bayesian" approach, in which case we refer to it as a "Bayesian conditional"). Treating conditional statements as material implications has been shown to be problematic [e.g., 1], hence the proposal to view them as Bayesian conditionals $\boldsymbol{P}(B \mid A)$ [e.g., 32]. At the same time, Bayesian conditionals are not appropriate for handling (indicative) conditionals [e.g., see 23, for a discussion], and cannot be used for inference (e.g., using *modus tollens*) if the probability of the prior is zero ($P(A) = 0$). Furthermore, various "triviality results" such as those by Lewis exposed difficulties associated with assigning conditional probabilities to conditional statements [24, 25] (see also [8]). And yet another difficulty with assigning probabilities is the "shiftiness" of conditionals, referring to the possibility that the same formal implication might have different interpretations when embedded in different contexts (see [22] for a detailed discussion).

Fortunately, there is a way to accommodate both probabilistic material implication and Bayesian conditionals under a common generalization, thus relieving some of the past tension between the two approaches and emphasizing their commonalities. Specifically, it is possible to define a one-parameter family of generalized implications, $\{A \xrightarrow{\gamma} B : \gamma \in [0, 1]\}$, for which the classical probabilistic implications correspond to the endpoints: $\gamma = 1$ gives probabilistic material implication $\boldsymbol{P}(A \to B)$, and $\gamma = 0$ gives the Bayesian conditional $\boldsymbol{P}(B \mid A)$, see [26].

The question then arises whether we need to reconceptualize (probabilistic) material implication and Bayesian conditionals in light of this generalization. The goal of this paper is to investigate some of the formal properties of the "generalized implication" $\xrightarrow{\gamma}$, in particular, the probability bounds arising from inferences involving $\xrightarrow{\gamma}$, as well as an interpretation of the γ-parameter. After a brief motivation for the generalization and the introduction of its mathematical form, we start with a review of its basic properties: functional relationships among various probabilities comparing material implication, Bayesian conditionals, and the generalized implication, as well as their bounds. We then present a result on bounds for generalized implication chains as they would occur in rule-based inference systems. Finally, we discuss a novel interpretation of the generalized implication and the formula that determines its probability, and also briefly discuss possible applications of generalized implications in logic and philosophy.

2. Motivation and Related Work

No consensus exists for the interpretation, modeling, and analysis of conditionals in natural language, but some popular approaches include formal logic and rules of inference, mental models which are prominent in the psychological literature, and suppositional theories that focus on estimating the probability of conditionals [5] (see [10], [36] for more discussion). In Artificial Intelligence research and applications the Bayesian conditional has become the *de facto* interpretation, as advocated by [32].

Efforts to combine probability and logic go back at least to Leibniz, Jakob Bernoulli, and Boole (e.g., see [18], and also [17] for a formal framework for "probability models"). Mapping logical operators onto set-theoretic operators in the usual way, we get set-theoretic, and hence, probabilistic interpretations of propositional sentences, e.g., from $A \to B \equiv \neg A \vee B \equiv \neg A \vee (A \wedge B)$ we get the standard

$$\boldsymbol{P}(A \to B) = \boldsymbol{P}(\overline{A}) + \boldsymbol{P}(A \cap B). \tag{1}$$

It was recognized at least as early as [2] (see also [17], [18]) that imprecise probabilities (for which we only have a lower and an upper bound) are needed when dealing with probabilities of events (i.e., sets representing propositional sentences). For our purposes, the most salient motivation for working with imprecise probabilities is that they are unavoidable if we want to use *modus ponens*, even if the implication rules are not probabilistic: applying *modus ponens* to infer $\boldsymbol{P}(B)$ only yields the bounds $\boldsymbol{P}(B) \in [\boldsymbol{P}(A \cap B),\ \boldsymbol{P}(A \to B)]$. Note that we can also use the "Bayesian implication" $B \,|\, A$ for inference rather than material implication to get bounds on B which were initially presented in [39] (see [34], [6], [9], and [12] for a development of the symbol "$B \,|\, A$", and see Section 8 for more on the relevance of [12] to this paper).

Using linear programming, the existence of lower and upper bound functions for the probability of any propositional logic formula was established in [16]. In [17], these results were extended, and probability bounds on *modus ponens* and *hypothetical syllogism* were given. A formal semantic method to compute the bounds on the probability of a given sentence in the predicate calculus was developed in [28] and extended in [11]. Probability bounds on *hypothetical syllogism* and other inference rules from propositional classical logic were computed in [21].

More generally, a formal proof system is needed for performing inference. An early proposal for probabilistic proof system was [28], and this was generalized in [13] by incorporating conditional probabilities. A complete proof system for probability logic was put forward by [31]. A Gentzen-style proof system was introduced in [3], [4]. For a discussion on the extension of probabilistic propositional logic to modal logics see [19]. An overview of probabilistic logics can be found in [15] (see also [30]).

A parametrized family of abstract probabilistic logical implications $A \xrightarrow{\gamma} B$ was introduced in [26], and the probability of such a generalized implication was computed as a function of the parameter γ, for $0 \leq \gamma \leq 1$:

$$\boldsymbol{P}(A \xrightarrow{\gamma} B) = \frac{\boldsymbol{P}(A \cap B) + \gamma \boldsymbol{P}(\overline{A})}{\boldsymbol{P}(A) + \gamma \boldsymbol{P}(\overline{A})}, \tag{2}$$

This is slightly different from the formula presented in [26], but this version more clearly shows the form to be some sort of interpolation from considering none of the probability of $\overline{A}$ to all of $\overline{A}$.

It is immediately apparent that $\gamma = 0$ corresponds to the Bayesian conditional, and $\gamma = 1$ corresponds to (the probability of) material implication, hence γ is interpolating between the Bayesian conditional and material implication. Hence, these two interpretations of the phrase "if A then B" are actually different manifestations of the same abstract concept—the two endpoints of a one-parameter family of probabilistic logical implications.

In this paper, we extend the results of [21] to include bounds on the generalized implication of [26], including *modus ponens*, *hypothetical syllogism*, and *modus tollens* with generalized implications. In particular, this shows how to treat bounds involving material implication and Bayesian conditionals in a uniform manner.

3. Computing probability bounds

We start by reviewing basic techniques for computing probability bounds relating to probabilistic *modus ponens* and *hypothetical syllogism* (see [20] for details).

3.1. Background and notation

As noted above, a generalized notion of a probabilistic logical implication, $A \xrightarrow{\gamma} B$, was introduced in [26], and its probability was computed to be (2) as follows. Fix arbitrary events A, B and consider probability distributions $\boldsymbol{P}$ that are constant on the four basic regions determined by A, B. Then define a probabilistic *logical* operation to be a function $F(\boldsymbol{P})$ taking values in $[0, 1]$, define probabilistic *implication* operators to be those probabilistic logical operators F such that, for any probability distributions $\boldsymbol{P}, \boldsymbol{P}'$, we have (i) if $\boldsymbol{P}(A \cap B) = \boldsymbol{P}'(A \cap B)$ and $\boldsymbol{P}(A \cap \overline{B}) = \boldsymbol{P}'(A \cap \overline{B})$, then $F(\boldsymbol{P}) = F(\boldsymbol{P}')$, and (ii) $\boldsymbol{P}(A \cap \overline{B}) = 0 \Rightarrow F(\boldsymbol{P}) = 1$. Then the "*natural*" implication operators were defined as those being linear in $\boldsymbol{P}(A \cap B)$. And finally, $\boldsymbol{P}(A \cap B)$ was computed as a function of $a = \boldsymbol{P}(A)$ and $g = \boldsymbol{P}(A \xrightarrow{\gamma} B)$. Then it was shown that if this function is commutative (in a and g) then g must be of the form given by (2). For more details, see [26].

We start by investigating the relationships among the generalized implication $A \xrightarrow{\gamma} B$, the material implication $A \to B$, the Bayesian implication $B \mid A$, the premise A, and the intersection $A \cap B$. Henceforth, we will use the following notational conventions:

Definition 1 (Notation for basic probability variables)**.** Fix events A, B, and $\gamma \in [0, 1]$. Let $X = A \cap B$. Then let $a = \boldsymbol{P}(A)$, $b = \boldsymbol{P}(B)$, $x = \boldsymbol{P}(X)$, $m = \boldsymbol{P}(A \to B)$ for *m*aterial implication, $c = \boldsymbol{P}(B \mid A)$ for Bayesian *c*onditional, and $g = \boldsymbol{P}(A \xrightarrow{\gamma} B)$ for *g*eneralized implication.

We can use the notation from Definition 1 to express the four basic regions in terms of a, b, m, c as $\boldsymbol{P}(A \cap B) = ac = m + a - 1$, $\boldsymbol{P}(A \cap \overline{B}) = a(1 - c) = 1 - m$, $\boldsymbol{P}(\overline{A} \cap B) = b - ac = 1 - m - a + b$, $\boldsymbol{P}(\overline{A} \cap \overline{B}) = 1 - a - b + ac = m - b$. Note that if $a = 0$ then $x = \boldsymbol{P}(A \cap B) = 0$, even though c is undefined, and the expressions are still correct using this interpretation of ac (as the intersection). But typically, we work under the assumption that $a \neq 0$, so that c is defined. Furthermore, since a, b, x, m, c, g are probabilities, we obviously must have $0 \leq a, b, x, m, c, g \leq 1$. Furthermore, we will also only consider $0 \leq \gamma \leq 1$ in this paper.

As discussed above, probability bounds are unavoidable, so all probabilities will henceforth be *imprecise*, i.e., we will typically not know the value p of a probability, but rather, will only be given $p_0, p_1 \in [0, 1]$ such that $p \in [p_0, p_1]$. The quantity p_0 is the *lower*, and p_1 is the *upper* probability for p. Of course, there is always the need to restrict to the interval $[0, 1]$ for the output of any formula giving probability bounds. In order to simplify notation, we will build this into our notation for probability bounds.

Definition 2 (The Lower- and Upper-bound operators)**.** Suppose that p_0, p_1 are lower and upper bounds on some probability $p \in [0, 1]$, i.e., $p_0 \leq p \leq p_1$, but suppose further that p_0, p_1 are computed from functions that do not necessarily give values in $[0, 1]$. Then

$$p_{\boldsymbol{L}} = \max\{p_0, 0\} \ , \quad p_{\boldsymbol{U}} = \min\{p_1, 1\}.$$

In what follows, we refer to this process as *clipping* (to keep lower and upper probability bounds in $[0, 1]$). For any probability p, we will write $p \in [p_{\boldsymbol{L}},\ p_{\boldsymbol{U}}]$, and will denote the width of any probability interval $[p_{\boldsymbol{L}},\ p_{\boldsymbol{U}}]$ as $\Delta p = p_{\boldsymbol{U}} - p_{\boldsymbol{L}}$.

3.2. Functional dependencies

We first observe that the quantities m, c, a, x, g, and γ can all be expressed in terms of each other using (2). For example, the standard relationship, given by (1), i.e., $m = 1 - a + x$, is easily recovered by letting $\gamma = 1$ in (2). Then we get the following functional relationships: $g(\gamma, a, x) = \dfrac{x + \gamma(1-a)}{a + \gamma(1-a)}$, $g(\gamma, a, c) = \dfrac{ac + \gamma(1-a)}{a + \gamma(1-a)}$, $g(\gamma, a, m) = \dfrac{(a + m - 1) + \gamma(1-a)}{(1-\gamma)a + \gamma \cdot 1}$, and in general, for arbitrary $\hat{\gamma}$ and its corresponding $\hat{g} = \boldsymbol{P}(A \xrightarrow{\hat{\gamma}} B)$, we have

$$g(\gamma, a, \hat{g}, \hat{\gamma}) = \frac{[(1-\hat{\gamma})a\hat{g} + \hat{\gamma}(a + \hat{g} - 1)] + \gamma(1-a)}{a + \gamma(1-a)}. \tag{3}$$

Other functional relationships are easily derived, e.g., expressions for $x(\gamma, a, g)$, $m(\gamma, a, g)$, $c(\gamma, a, g)$, $a(\gamma, g, c)$, etc.

3.3. Using partial derivatives to compute bounds

Then these functional relationships were used to compute probability bounds. A simple example will suffice to explain the idea. Fix $\gamma > 0$ (for the sake of simplicity). To compute the bounds $g_{\boldsymbol{L}}, g_{\boldsymbol{U}}$ for g given the bounds $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, $c \in [c_{\boldsymbol{L}},\ c_{\boldsymbol{U}}]$, we must consider the function $g(\gamma, a, c) = \dfrac{ac + \gamma(1-a)}{a + \gamma(1-a)}$, and simply compute the partial derivatives $\dfrac{\partial g}{\partial a} = \dfrac{-\gamma(1-c)}{(a + \gamma(1-a))^2} \leq 0$ and $\dfrac{\partial g}{\partial c} =$

$\dfrac{a}{a+\gamma(1-a)} \geq 0$. From these partial derivatives we see that the function $\lambda a g(\gamma, a, c)$ is monotonic decreasing on $[0,1]$ for every $0 \leq c < 1$, and similarly, $\lambda c g(\gamma, a, c)$ is monotonic increasing. Hence, the max and min, i.e., the bounds, occur at the corresponding corners, i.e., $g_{\boldsymbol{L}} = g(\gamma, a_{\boldsymbol{U}}, c_{\boldsymbol{L}})$, $g_{\boldsymbol{U}} = g(\gamma, a_{\boldsymbol{L}}, c_{\boldsymbol{U}})$.

4. *Modus Ponens* for generalized implications (bounds on b)

In performing logical inference, *modus ponens*, of course, plays a key role. In the probabilistic context, *modus ponens* corresponds to computing (bounds on) $b = \boldsymbol{P}(B)$ given (bounds on) $a = \boldsymbol{P}(A)$ and (bounds on) $g = \boldsymbol{P}(A \xrightarrow{\gamma} B)$. The term *modus ponens* is usually reserved for the case when $\gamma = 1$ (so $g = m$). If $\gamma = 0$, we have Bayesian *modus ponens*. Of course, now we want bounds on b when γ is arbitrary.

The equations in Section 3.2 are true for any B, but b is itself not functionally related to any of g, a, x. Consequently, we can only get bounds on b, even if we know all of g, a, x precisely. In this case, we use $A \cap B \subseteq B \subseteq \overline{A} \cup B$, to get the well-known (tight) bounds $x \leq b \leq m$, i.e., $b_{\boldsymbol{L}} = x$, $b_{\boldsymbol{U}} = m$.

Simple algebraic manipulation of the equations in Section 3.2 allows us to express these bounds in terms of a, γ, g:

$$b_{\boldsymbol{L}} = (1-\gamma)ag + \gamma(a+g-1), \tag{4}$$
$$b_{\boldsymbol{U}} = (1-\gamma)ag + \gamma(a+g-1) + (1-a). \tag{5}$$

So b does not depend on a, γ, g in a functional way, but its lower and upper bounds do, i.e., for fixed γ we could have written (4), (5) as $b_{\boldsymbol{L}}(a,g) = \cdots$ and $b_{\boldsymbol{U}}(a,g) = \cdots$. Of course, these formulas apply when we have precise probabilities for a, g. As in Section 3.3, if we only have bounds for a and g, we compute $\dfrac{\partial}{\partial a} b_{\boldsymbol{L}}(a,g) > 0$, $\dfrac{\partial}{\partial g} b_{\boldsymbol{L}}(a,g) > 0$, and $\dfrac{\partial}{\partial a} b_{\boldsymbol{U}}(a,g) < 0$, $\dfrac{\partial}{\partial g} b_{\boldsymbol{U}}(a,g) > 0$. Recall that $x = \boldsymbol{P}(A \cap B)$. So from (4) we can compute $(b_{\boldsymbol{L}})_{\boldsymbol{L}} = x_{\boldsymbol{L}}$, $(b_{\boldsymbol{L}})_{\boldsymbol{U}} = x_{\boldsymbol{U}}$, and from (5) we compute $(b_{\boldsymbol{U}})_{\boldsymbol{L}} = m_{\boldsymbol{L}}$, $(b_{\boldsymbol{U}})_{\boldsymbol{U}} = m_{\boldsymbol{U}}$.

We will abuse notation and write $b_{\boldsymbol{L}} = (b_{\boldsymbol{L}})_{\boldsymbol{L}}$ and $b_{\boldsymbol{U}} = (b_{\boldsymbol{U}})_{\boldsymbol{U}}$, and discard the other two, i.e., $b_{\boldsymbol{L}}$ refers to both the *value* of the bound (computed from other bounds), as well as the *function* that is used to compute this bound. This leads to

$$b_{\boldsymbol{L}} = b_{\boldsymbol{L}}(a_{\boldsymbol{L}}, g_{\boldsymbol{L}}) = (1-\gamma)a_{\boldsymbol{L}}g_{\boldsymbol{L}} + \gamma(a_{\boldsymbol{L}} + g_{\boldsymbol{L}} - 1) \tag{6}$$
$$b_{\boldsymbol{U}} = b_{\boldsymbol{U}}(a_{\boldsymbol{L}}, g_{\boldsymbol{U}}) = (1-\gamma)a_{\boldsymbol{L}}g_{\boldsymbol{U}} + \gamma(a_{\boldsymbol{L}} + g_{\boldsymbol{U}} - 1) + (1-a_{\boldsymbol{L}}). \tag{7}$$

Whether $b_{\boldsymbol{L}}$ refers to the function in (4) or the value in (6) will be clear from context. Furthermore, $\Delta b = (1-a_{\boldsymbol{L}}) + [(1-\gamma)a_{\boldsymbol{L}} + \gamma \cdot 1]\Delta g$, and if g is precise we just have $\Delta b = 1 - a_{\boldsymbol{L}}$.

5. Generalized hypothetical syllogism

Implications chains are of particular interest for rule-based reasoning. By an implication chain we mean a sequence of generalized implications $B^i \xrightarrow{\gamma^i} B^{i+1}$, for $i = 0, 1, \ldots, N-1$. Let $A = B^0$, and suppose that we are the given bounds $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$ for the initial antecedent A.

There are two flavors of inference rule involving implication chains, both referred to as *hypothetical syllogism*. The first is a generalization of *modus ponens*: assume $B^0,\ B^0 \to B^1,\ B^1 \to B^2$, conclude B^2. The second is of the form: assume $B^0 \to B^1,\ B^1 \to B^2$, conclude $B^0 \to B^2$. These are, of course, easily generalized to more than two hypothetical premises. We will, of course, be dealing with the probabilized versions of these. The results of this section are from [20].

5.1. Flavor I: one categorical premise, multiple hypothetical premises

We want to compute the bounds for the final consequent b^N. We first do this for arbitrary γ. Then we easily derive all the corresponding classical bounds by setting $\gamma = 0, 1$.

Theorem 3. *Fix arbitrary γ and suppose we are given a sequence of generalized implications $B^i \xrightarrow{\gamma} B^{i+1}$, for $i = 0, 1, \ldots, N-1$. Suppose $A = B^0$ and that we are given bounds $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, $g^i \in [g^i_{\boldsymbol{L}},\ g^i_{\boldsymbol{U}}]$, for $i < N$. Let $y^i = \gamma(1 - g^i) + g^i$ and $z^i = \gamma(1 - g^i)$. Then the bounds at the end of the generalized implication chain are*

$$b^N_{\boldsymbol{L}} = a_{\boldsymbol{L}} \prod_{i=0}^{N-1} y^i_{\boldsymbol{L}} - \sum_{i=0}^{N-1} \left(z^i_{\boldsymbol{U}} \prod_{j=i+1}^{N-1} y^j_{\boldsymbol{L}} \right),$$

$$b^N_{\boldsymbol{U}} = 1 - z^{N-1}_{\boldsymbol{L}} \ - \ (1 - y^{N-1}_{\boldsymbol{U}}) \left[a_{\boldsymbol{L}} \prod_{i=0}^{N-2} y^i_{\boldsymbol{U}} - \sum_{i=0}^{N-2} \left(z^i_{\boldsymbol{L}} \prod_{j=i+1}^{N-2} y^j_{\boldsymbol{U}} \right) \right].$$

Note that since $\gamma = 0 \Rightarrow y = c,\ z = 0$, and $\gamma = 1 \Rightarrow y = 1,\ z = 1 - m$, we can easily recover the classical cases below. However, we do not ascribe any significance to y, z other than that they allow the expressions to be written more compactly.

Corollary 4. If we are using material implications in our chain, we just set $\gamma = 1$ in the formulas above. So $m^i \in [m^i_{\boldsymbol{L}},\ m^i_{\boldsymbol{U}}]$, for $i < N$, are given, and the bounds at the end of the implication chain are

$$b^N_{\boldsymbol{L}} = a_{\boldsymbol{L}} - \sum_{i=0}^{N-1} \left(1 - m^i_{\boldsymbol{L}}\right) \ , \qquad b^N_{\boldsymbol{U}} = m^{N-1}_{\boldsymbol{U}}.$$

We note that for typical data (not all $m^i_{\boldsymbol{L}} = 1$) we have $b^N_{\boldsymbol{L}} \to 0$. Also, $b^N_{\boldsymbol{U}}$ just depends on the last implication.

Corollary 5. For Bayesian *modus ponens*, we have $\gamma = 0$, and $c^i \in [c^i_{\boldsymbol{L}},\ c^i_{\boldsymbol{U}}]$, for $i < N$ are given. Then the bounds at the end of the implication chain are

$$b^N_{\boldsymbol{L}} = a_{\boldsymbol{L}} \prod_{i=0}^{N-1} c^i_{\boldsymbol{L}}\ , \quad b^N_{\boldsymbol{U}} = 1 - \left(a_{\boldsymbol{L}} \prod_{i=0}^{N-1} c^i_{\boldsymbol{L}} \right) \left(1 - c^{N-1}_{\boldsymbol{U}}\right) .$$

We note that for typical data (not all $c^i_{\boldsymbol{L}} = 1$) we see $b^N_{\boldsymbol{L}} \to 0$ and $b^N_{\boldsymbol{U}} \to 1$.

Next we discuss the most general case, which generalizes Theorem 3 above. This is intractable on its own, but can be approximated using Theorem 3.

Lemma 6. *Let $B^i \xrightarrow{\gamma^i} B^{i+1}$, for $i = 0, 1, \ldots, N-1$, be a given chain of generalized implications. Let $A = B^0$, and suppose we are given the bounds $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, and the bounds $g^i \in [g^i_{\boldsymbol{L}},\ g^i_{\boldsymbol{U}}]$, for $i < N$. Then bounds for b^N the end of the generalized implication chain can be computed using Theorem 3.*

5.2. Flavor II: only hypothetical premises

Lemma 7. *A generalized implication chain, $B^i \xrightarrow{\gamma} B^{i+1}$, for $i = 0, 1, \ldots, N-1$, from Theorem 3, can be replaced with a single implication $A \xrightarrow{\gamma} B$, where $A = B^0$, $B = B^N$, and the bounds $[b^N_{\boldsymbol{L}},\ b^N_{\boldsymbol{U}}]$ are the same, whether computed from the implication chain or the single implication.*

Lemma 8. *Let $B^i \xrightarrow{\gamma^i} B^{i+1}$, for $i = 0, 1, \ldots, N-1$, be a generalized implication chain from Lemma 6, and let γ be arbitrary. Then this implication chain can be replaced with a single implication $A \xrightarrow{\gamma} B$, where $A = B^0$, $B = B^N$, and the bounds $[b^N_{\boldsymbol{L}},\ b^N_{\boldsymbol{U}}]$ are the same, whether computed from the implication chain or the single implication (when these both use γ).*

Several consequences of this result are noteworthy:

(a) The upper bound $a_{\boldsymbol{U}}$ of the initial antecedent A plays no role in the determination of the bounding interval of the final consequent B^N. This comes as no surprise because the same is true for a single rule (see (7)).

(b) The lower bound $b^N_{\boldsymbol{L}}$ of the consequent B^N is completely determined by the lower bounds of the antecedent A and the rules being cascaded together; their upper bounds play no role in its determination. In fact, if the lower bound of the antecedent A or any of the rules is zero, the lower bound of the consequent is zero.

(c) In the case $\gamma = 1$ (material implication) the upper bound $b^N_{\boldsymbol{U}}$ of the consequent B^N is completely determined by, in fact it is equal to, the upper bound of the last rule $B^{N-1} \to B^N$; neither its lower bound nor the antecedent A nor the other rules play a role in the determination of $b^N_{\boldsymbol{U}}$.

(d) In the case $\gamma = 1$ (material implication), we compute $\Delta b^N = b^N_{\boldsymbol{U}} - b^N_{\boldsymbol{L}} = \Delta b^N = \Delta m^{N-1} + (1 - a_{\boldsymbol{L}}) + \sum_{i=0}^{N-2}(1 - m^i_{\boldsymbol{L}})$. Interestingly, this implies that $\Delta b^N \geq \Delta m^{N-1}$, i.e., the uncertainty interval associated with the consequent

B^N can never be narrower than the uncertainty interval associated with the rule at the end, viz., $B^{N-1} \to B^N$. In fact,

$$\Delta b^N \begin{cases} = \Delta m^{N-1}, & \text{if } m_{\boldsymbol{L}}^{N-1} = 0 \text{ or} \\ & \quad a_{\boldsymbol{L}} = m_{\boldsymbol{L}}^i = 1,\ 0 \le i \le n-2; \\ > \Delta m^{N-1}, & \text{otherwise.} \end{cases}$$

6. Converses

In this section we will consider the logical converse $B \xrightarrow{\gamma} A$ of our original sentence $A \xrightarrow{\gamma} B$. As usual, fix γ and establish the notation $k = \boldsymbol{P}(B \xrightarrow{\gamma} A)$. Our interest lies in trying to express k in terms of $[a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, $[g_{\boldsymbol{L}},\ g_{\boldsymbol{U}}]$ from the original sentence.

6.1. The classical case

For the case of material implication, i.e., $\gamma = 1$, we trivially have the converse probability $k = \boldsymbol{P}(B \to A) = 1 - b + x$, As in Section 3.3 we get the bounds $k_{\boldsymbol{L}} = k(b_{\boldsymbol{U}}, x_{\boldsymbol{L}})$ and $k_{\boldsymbol{U}} = k(b_{\boldsymbol{L}}, x_{\boldsymbol{U}})$. For the sake of clarification, here is an easy *wrong* calculation: $k_{\boldsymbol{L}} = k(b_{\boldsymbol{U}}, x_{\boldsymbol{L}}) = 1 - b_{\boldsymbol{U}} + x_{\boldsymbol{L}} = 1 - m_{\boldsymbol{U}} + (a_{\boldsymbol{L}} + m_{\boldsymbol{L}} - 1) = a_{\boldsymbol{L}} - \Delta m$. First, this formula does not always simplify to $k_{\boldsymbol{L}} = a_{\boldsymbol{L}} - \Delta m$ because of Definition 2, eg., if $a_{\boldsymbol{L}} = m_{\boldsymbol{L}} = 0$, the quantity in the parentheses $x_{\boldsymbol{L}}$ is 0, leaving $k_{\boldsymbol{L}} = 1 - m_{\boldsymbol{U}} \ge 0$. But if we only apply the restriction to $[0, 1]$ at the end of the computation, we would get $k_{\boldsymbol{L}} = 0$. A second, more serious, problem is that k is not functionally related to a, m since b is not, so there is no function $k(a, m)$, to which we can apply the techniques of Section 3.3 to compute $k_{\boldsymbol{L}}$ (in terms of a, m). So in the equation above we have $b_{\boldsymbol{U}} = m \le m_{\boldsymbol{U}}$, but the inequality will be strict unless m is known precisely.

The point is that $b_{\boldsymbol{L}}, b_{\boldsymbol{U}}$ *are* functionally related to a and m, even though b is not. So the solution is to introduce a new variable and functional relationship, $k_l = k(b_{\boldsymbol{U}}(a, m), x(a, m))$, for the lower bound of k, and then compute $(k_l)_{\boldsymbol{L}}$ as in Section 3.3. Then we set $k_{\boldsymbol{L}} = (k_l)_{\boldsymbol{L}}$ (see the discussion around (6)). We get $k_l = 1 - m + (a + m - 1) = 1 - m + x = a$, so $(k_l)_{\boldsymbol{L}} = a_{\boldsymbol{L}} = (1 - m_{\boldsymbol{U}})/(1 - c_{\boldsymbol{L}})$. A similar calculation with the new variable k_u for the upper bound ultimately gives $[k_{\boldsymbol{L}},\ k_{\boldsymbol{U}}] = [a_{\boldsymbol{L}},\ 1] = [(1 - m_{\boldsymbol{U}})/(1 - c_{\boldsymbol{L}}),\ 1]$. Hence, if $c_{\boldsymbol{L}} = 0$ and m is known precisely, then we can only conclude $[k_{\boldsymbol{L}},\ k_{\boldsymbol{U}}] = [1 - m, 1]$.

For the case of the Bayesian conditional, $\gamma = 0$, the converse of $\boldsymbol{P}(B \mid A)$ would be $\boldsymbol{P}(A \mid B)$. Of course, this just requires Bayes' formula, and $k = \boldsymbol{P}(A \mid B) = ac/b$, so $k_{\boldsymbol{L}} = a_{\boldsymbol{L}} c_{\boldsymbol{L}} / b_{\boldsymbol{U}}$, $k_{\boldsymbol{U}} = a_{\boldsymbol{U}} c_{\boldsymbol{U}} / b_{\boldsymbol{L}}$. The analysis is similar to the material case, and leads to $[k_{\boldsymbol{L}},\ k_{\boldsymbol{U}}] = [a_{\boldsymbol{L}} c_{\boldsymbol{L}} / (1 - a_{\boldsymbol{L}} + a_{\boldsymbol{L}} c_{\boldsymbol{L}}),\ 1]$. It is worth noting that the denominator in $k_{\boldsymbol{L}}$ is not $m_{\boldsymbol{L}} = 1 - a_{\boldsymbol{U}} + a_{\boldsymbol{U}} c_{\boldsymbol{L}}$ (unless a is precise). Hence, if $a \in (0, 1]$ and $c > 0$ is known precisely, then we can only conclude $[k_{\boldsymbol{L}},\ k_{\boldsymbol{U}}] = (0, 1]$. (Of course, since we need $a > 0$ for c to be defined we don't write $k \in [0, 1]$.)

6.2. Converses for arbitrary $\boldsymbol{\gamma}$

Applying the relevant equation of Section 3.2 to $\boldsymbol{P}(B \xrightarrow{\gamma} A)$ (but substituting k for g and b for a) gives $k = \dfrac{x+\gamma(1-b)}{b+\gamma(1-b)} = \dfrac{ac+\gamma(1-b)}{b+\gamma(1-b)}$. Then, as in Section 3.3, we have $k_{\boldsymbol{L}} = k(a_{\boldsymbol{L}}, c_{\boldsymbol{L}}, b_{\boldsymbol{U}})$, $k_{\boldsymbol{U}} = k(a_{\boldsymbol{U}}, c_{\boldsymbol{U}}, b_{\boldsymbol{L}})$, since $\partial k/\partial a > 0$, $\partial k/\partial c > 0$, $\partial k/\partial b < 0$. As above, we define new variables and functional relationships $k_l = k(a, c, b_{\boldsymbol{U}}(a, c))$ and $k_u = k(a, c, b_{\boldsymbol{L}}(a, c))$. Since $b_{\boldsymbol{U}} = m$ we have $k_l = k(a, c, m_{\boldsymbol{U}}(a, c)) = \dfrac{ac+\gamma a(1-c)}{1-a+ac+\gamma a(1-c)}$, and as in Section 3.3 we get $(k_l)_{\boldsymbol{L}}(a_{\boldsymbol{L}}, c_{\boldsymbol{L}})$. Similarly, $b_{\boldsymbol{L}} = x$ gives $k_u = k(a, c, x(a, c)) = \dfrac{ac+\gamma(1-ac)}{ac+\gamma(1-ac)} = 1$, and hence,

$$k_{\boldsymbol{L}} = \frac{a_{\boldsymbol{L}}c_{\boldsymbol{L}} + \gamma a_{\boldsymbol{L}}(1-c_{\boldsymbol{L}})}{1-a_{\boldsymbol{L}}+a_{\boldsymbol{L}}c_{\boldsymbol{L}}+\gamma a_{\boldsymbol{L}}(1-c_{\boldsymbol{L}})}\ , \qquad k_{\boldsymbol{U}} = 1.$$

It is interesting to note that the alternate expression $k_l = \dfrac{ac+\gamma(1-m)}{m+\gamma(1-m)}$, yielding $k_{\boldsymbol{L}} = \dfrac{a_{\boldsymbol{L}}c_{\boldsymbol{L}}+\gamma(1-m_{\boldsymbol{U}})}{m_{\boldsymbol{U}}+\gamma(1-m_{\boldsymbol{U}})}$, gives an inferior bound unless c is precise.

7. *Modus Tollens* (contrapositives)

Suppose we have bounds on $m = \boldsymbol{P}(A \to B)$. In classical logic, if we happen also to know bounds on B, and would like to transfer that knowledge into bounds on A, we need the contrapositive. In this section we will consider the contrapositive $\overline{B} \xrightarrow{\gamma} \overline{A}$ of our original generalized implication $A \xrightarrow{\gamma} B$. Note that γ is the same in both of these implications, but the probabilities of each sentence will typically be different. Fix γ and establish the notation $q = \boldsymbol{P}(\overline{B} \xrightarrow{\gamma} \overline{A})$.

7.1. Finding bounds on a from bounds on b

If we have no information on a, i.e., $a \in [0,1]$, but we have bounds on g and b (hence $\overline{b}$), and we know γ precisely, we want to find (better) bounds on a.

Solving (4) for a gives $a \le (b+\gamma(1-g))/(g+\gamma(1-g))$, and to find the worst case of this upper bound we compute the relevant derivatives to get $a \le \dfrac{b_{\boldsymbol{U}}+\gamma(1-g_{\boldsymbol{L}})}{g_{\boldsymbol{L}}+\gamma(1-g_{\boldsymbol{L}})}$. Then $\gamma = 0 \Rightarrow a \le \dfrac{b_{\boldsymbol{U}}}{c_{\boldsymbol{L}}}$, and $\gamma = 1 \Rightarrow a \le 1 - m_{\boldsymbol{L}} + b_{\boldsymbol{U}}$. Similarly, solving (5) for a gives $a \le (1-b+\gamma(1-g))/((1-\gamma)(1-g))$. Then, using the technique of Section 3.3, we get $a \le \dfrac{1-b_{\boldsymbol{L}}+\gamma(1-g_{\boldsymbol{U}})}{(1-\gamma)(1-g_{\boldsymbol{U}})}$. Hence, $\gamma = 0 \Rightarrow a \le (1-b_{\boldsymbol{L}})/(1-c_{\boldsymbol{U}})$, and for $\gamma = 1$ the formula is not defined, but yields the trivial bound $a \le 1$ since the formula is unbounded as $\gamma \to 1$ (and we use Definition 2). More interestingly, (4) and (5), the lower and upper bounds formulas for b, both gave upper bounds for a. So at this point we have

$$a_{\boldsymbol{U}} = \min\left\{ \frac{b_{\boldsymbol{U}} + \gamma(1-g_{\boldsymbol{L}})}{g_{\boldsymbol{L}} + \gamma(1-g_{\boldsymbol{L}})} \, , \, \frac{1 - b_{\boldsymbol{L}} + \gamma(1-g_{\boldsymbol{U}})}{(1-\gamma)(1-g_{\boldsymbol{U}})} \right\}.$$

These bounds, for the special case of $\gamma = 0$ ($g = c$) and precise quantities, were initially reported in [39].

If we apply (4) to $\overline{B} \xrightarrow{\gamma} \overline{A}$ by substituting $\bar{b}, \bar{a}$ for a, b (resp.), then we must also use q (which we typically don't have) rather than g. The only time we have q is when $\gamma = 1$, in which case $q = g = m$. Since the classical bounds ($\gamma = 1$, $g = m$) are well-known, we won't pursue this approach here. However, we can express the inequality in (5) as applied to $\overline{B} \xrightarrow{\gamma} \overline{A}$, but in terms of γ, a, b, x from the start. These are the quantities that we have (as opposed to q). So $\bar{a} \le \boldsymbol{P}(\overline{A} \cap \overline{B}) + \boldsymbol{P}(\overline{\overline{B}}) \Rightarrow 1-a \le 1-a-b+x+b \Rightarrow 0 \le x = (1-\gamma)ag + \gamma(a+g-1)$, and solving for a gives $a \ge (\gamma(1-g))/(g + \gamma(1-g))$. Since this is the only expression of an upper bound for a that does not involve q, and $\partial/\partial g < 0$, the technique from Section 3.3 gives

$$a_{\boldsymbol{L}} = \frac{\gamma(1-g_{\boldsymbol{U}})}{g_{\boldsymbol{U}} + \gamma(1-g_{\boldsymbol{U}})}.$$

Then $\gamma = 0 \Rightarrow a_{\boldsymbol{L}} = 0$, and $\gamma = 1 \Rightarrow a_{\boldsymbol{L}} = 1 - m_{\boldsymbol{U}}$.

7.2. Bounds on q and the effect on modus tollens

As usual, we fix γ and consider $A \xrightarrow{\gamma} B$. As above, let $q = \boldsymbol{P}(\overline{B} \xrightarrow{\gamma} \overline{A})$. Then, as in Section 3.2 (after substituting q for g, $1-b$ for a, and $\boldsymbol{P}(\overline{A} \cap \overline{B}) = 1-a-b+x$ for x), we get $q = \dfrac{(1-a-b+x) + \gamma(1-(1-b))}{(1-b) + \gamma(1-(1-b))} = \dfrac{m - (1-\gamma)b}{1-(1-\gamma)b}$. If we have no further information on the bounds of b, we just use $x \le b \le m$, and $\partial q/\partial b < 0$, to get $q_{\boldsymbol{L}} = \dfrac{\gamma m}{1-(1-\gamma)m}$, $q_{\boldsymbol{U}} = \dfrac{m-(1-\gamma)x}{1-(1-\gamma)x}$. Of course, Δq measures how well *modus tollens* works for a particular γ. If $\gamma = 1$, then $\Delta q(m, x) = 0$, and the general case is

$$\Delta q = \frac{(1-\gamma)(1-m)(m-x)}{(1-(1-\gamma)m)(1-(1-\gamma)x)}.$$

Since $\partial \Delta q/\partial \gamma \le 0$ (for all m, x), this is a decreasing function on $[0, 1]$, with max at $\gamma = 0$, taking the value $(m-x)/(1-x)$. Figure 1 is a visual comparison of $\Delta q(m, x)$ for various γ. The value of γ does not have to be too large for the degradation of Δq to become relatively small. Of course, the specific maximum value of $\Delta q(m, x)$ could be computed as a function of γ, as well as the average degradation, which could be computed by an integral giving the volume under the surface: $\Delta q_{\text{avg}} = \frac{1}{2}\int_0^1 \int_0^m \Delta q(\gamma; x, m)\, dx\, dm$. Note: The inconsistent region comes from $m < x$, i.e., $a > 1$.

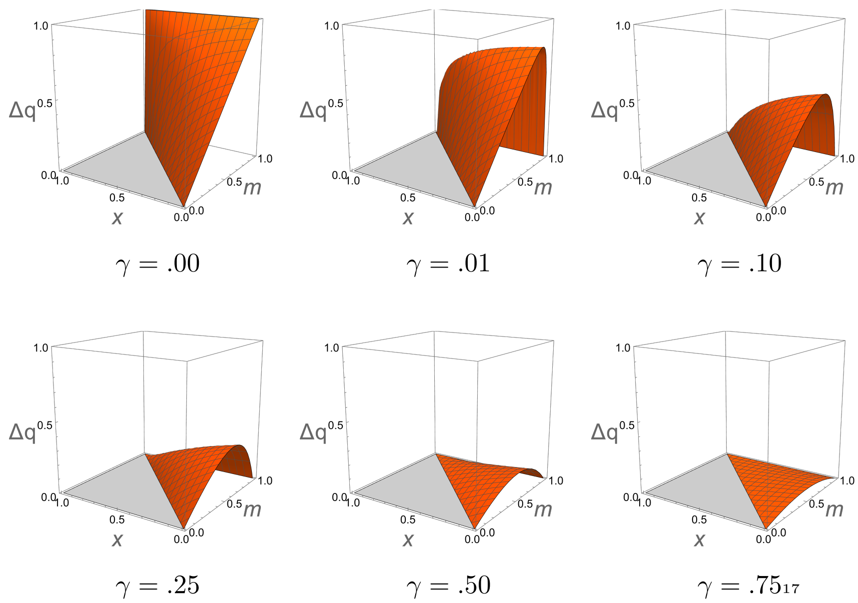


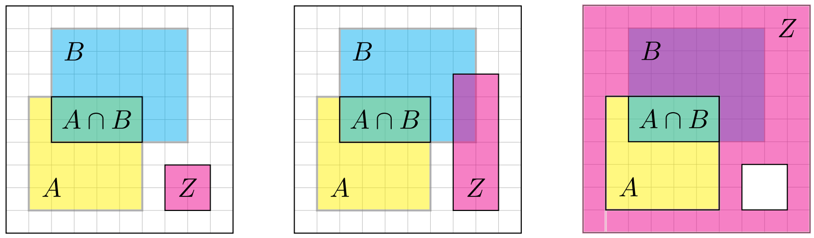
Figure 1: Δq measures the degradation of *modus tollens*.

8. Interpretation of γ

One question we have so far not addressed is how to interpret γ other than stating that the limit cases of $\gamma = 0$ and $\gamma = 1$ correspond to the Bayesian conditional and the material implication. In this section we will see how both $g = \boldsymbol{P}(A \xrightarrow{\gamma} B)$ and γ can be viewed as conditional probabilities. We begin by considering the easy equivalence $A \to (A \to B) \equiv A \to B$. The left-hand side of this equivalence is of the form $A \to D$, and hence corresponds to $\boldsymbol{P}(D \,|\, A) = \boldsymbol{P}(A \to B \,|\, A)$ in a Bayesian interpretation. Of course, the right-hand-side corresponds to $\boldsymbol{P}(B \,|\, A)$. Then the above logical equivalence suggests the easily verified equation,

$$\boldsymbol{P}(A \to B \,|\, A) = \boldsymbol{P}(B \,|\, A). \tag{8}$$

We note that Figure 2 indicates how to view (2) in a Venn diagram, and that it anticipates sections 8.1, 8.2 below.



(a) $\gamma = \frac{4}{75}$ (b) $\gamma = \frac{12}{75}$ (c) $\gamma = \frac{71}{75}$

Figure 2: $\boldsymbol{P}(A) = 0.25$, $\boldsymbol{P}(B) = 0.30$, $\boldsymbol{P}(A \cap B) = 0.08$, so we have $\boldsymbol{P}(A \xrightarrow{\mathbf{0}} B) = \boldsymbol{P}(B \,|\, A) = \frac{0.08}{0.25} = 0.32$ and $\boldsymbol{P}(A \xrightarrow{\mathbf{1}} B) = \boldsymbol{P}(A \to B) = 0.08 + 0.75 = 0.83$. Then, using $\boldsymbol{P}(Z) = \gamma \boldsymbol{P}(\overline{A})$ we get (a) $\boldsymbol{P}(Z) = 0.4 \Rightarrow \gamma = \frac{0.04}{0.75} = 0.053$, and hence, $\boldsymbol{P}(A \xrightarrow{\mathbf{0.053}} B) = \frac{0.08+0.04}{0.25+0.04} = 0.414$, (b) $\boldsymbol{P}(Z) = 0.12 \Rightarrow \gamma = \frac{0.12}{0.75} = 0.160$, and hence, $\boldsymbol{P}(A \xrightarrow{\mathbf{0.160}} B) = \frac{0.08+0.12}{0.25+0.12} = 0.541$, and (c) $\boldsymbol{P}(Z) = 0.71 \Rightarrow \gamma = \frac{0.71}{0.75} = 0.947$, and hence, $\boldsymbol{P}(A \xrightarrow{\mathbf{0.947}} B) = \frac{0.08+0.71}{0.25+0.71} = 0.823$.

8.1. Viewing g as a conditional probability

We now observe that a change to the conditioning set in the left-hand side of (8) will lead to the generalized implication $A \xrightarrow{\gamma} B$. Let $A_\gamma = A \cup Z$, where Z is any set such that $Z \subseteq \overline{A}$ and $\boldsymbol{P}(Z) = \gamma \boldsymbol{P}(\overline{A}) = \gamma(1 - \boldsymbol{P}(A))$. Hence, $\boldsymbol{P}(A_\gamma) = \boldsymbol{P}(A) + \gamma \boldsymbol{P}(\overline{A})$. Then, using $A \cap B \subseteq A_\gamma$ and $\overline{A} \cap A_\gamma = Z$, we get

$$\begin{aligned}
\boldsymbol{P}(A \to B \,|\, A_\gamma) &= \frac{\boldsymbol{P}((\overline{A} \cup B) \cap A_\gamma)}{\boldsymbol{P}(A_\gamma)} \\
&= \frac{\boldsymbol{P}(Z \cup ((B \cap A) \cup (B \cap Z)))}{\boldsymbol{P}(A_\gamma)} \\
&= \frac{\boldsymbol{P}(A \cap B) + \boldsymbol{P}(Z)}{\boldsymbol{P}(A_\gamma)} \\
&= \frac{\boldsymbol{P}(A \cap B) + \gamma \boldsymbol{P}(\overline{A})}{\boldsymbol{P}(A) + \gamma \boldsymbol{P}(\overline{A})} = \boldsymbol{P}(A \xrightarrow{\gamma} B),
\end{aligned} \tag{9}$$

where the last equality is just (2). In other words, $\boldsymbol{P}(A \xrightarrow{\gamma} B)$ really measures the conditional probability of the material implication $A \to B$ given any set $A_\gamma \supseteq A$ with $\boldsymbol{P}(A_\gamma) = \boldsymbol{P}(A) + \gamma \boldsymbol{P}(\overline{A})$.

8.2. Viewing γ as a conditional probability

We now give a concrete example that leads to an interpretation of γ as a conditional probability as well. In the following we will assume that B is known precisely, but that there are measurement errors associated with A. First, we establish some notation.

Definition 9. Let $\widehat{A}$ be the measured approximation to A, let the set of false positives be $F^+ = \widehat{A} \cap \overline{A}$, and the set of false negatives be $F^- = \widehat{\overline{A}} \cap A$. We also let $F = F^- \cup F^+$. Define the false positive rate to be $r^+ = \boldsymbol{P}(F^+ \,|\, \overline{A}) = \boldsymbol{P}(\widehat{A} \,|\, \overline{A})$, and the false negative rate to be $r^- = \boldsymbol{P}(F^- \,|\, A) = \boldsymbol{P}(\widehat{\overline{A}} \,|\, A)$. Finally, let $A^+ = A \cup F^+$ and $A^- = A \smallsetminus F^-$.

Now the special case in which γ is a conditional probability is as follows. Assume $r^- = 0$, and recall the sets $Z \subseteq \overline{A}$ and $A_\gamma = A \cup Z$ from Section 8.1. Since $F^+ \subseteq \overline{A}$ and $\boldsymbol{P}(F^+) = r^+ \boldsymbol{P}(\overline{A})$, we can set $Z = F^+$ and $\gamma = r^+$, so that $A_\gamma = A_{r^+} = A \cup F^+ = \widehat{A} = A^+$. From $A_{r^+} = \widehat{A}$ and (9), we immediately get

$$\boldsymbol{P}(A \to B \,|\, \widehat{A}) = \boldsymbol{P}(A \xrightarrow{r^+} B). \tag{10}$$

An alternative to the approach leading to (10) would be to replace the antecedent of the material implication in (8) with an appropriately defined $A_\gamma \subseteq A$, rather than the conditioning set. This will lead to a different interpretation of g, γ as a conditional probabilities (see Case 2b below). The main point of the following discussion is to explore what the semantics of the generalized implication $A \xrightarrow{\gamma} B$ might be, and what might be expressed by such a sentence. We explore this by computing the conditional probabilities arising from $\boldsymbol{P}(A \to B \,|\, A)$ by replacing one or both occurrences of A with $\widehat{A}, A^+, A^-$, under the assumption of r^+, r^- being zero or non-zero.

We begin with some notation. To extend any decoration $\sigma \in \{\,\widehat{\ }\,,\ ^+,\ ^-\}$ from Definition 9 to other quantities, we just replace A with A^σ in their definition. This leads to the following.

Definition 10. Let $\widehat{X} = \widehat{A} \cap B$, $X^+ = A^+ \cap B$, $X^- = A^- \cap B$, and similarly, $\widehat{c} = \boldsymbol{P}(B \,|\, \widehat{A})$, $c^+ = \boldsymbol{P}(B \,|\, A^+)$, $c^- = \boldsymbol{P}(B \,|\, A^-)$.

In the computations below there are several more quantities that arise naturally. We collect these in the next definition.

Definition 11. Let the Bayesian converse of r^- be $\widetilde{r}^- = \boldsymbol{P}(A \,|\, \widehat{\overline{A}})$. Also, define $u^- = \boldsymbol{P}(F^- \,|\, \overline{A^-})$, $u^+ = \boldsymbol{P}(F^+ \,|\, \overline{A^-})$, and $u = u^- + u^+$.

Case 1. (False positives only)

Assume $r^- = 0$, set $\gamma = r^+$, so $A_\gamma = A_{r^+} = A \cup F^+ = \widehat{A} = A^+$.

a) $\boldsymbol{P}(A \to B \,|\, \widehat{A}) = \boldsymbol{P}(A \xrightarrow{r^+} B)$ is just (10) above. It is only listed again here for the sake of Table 1.

b) Next we use $\widehat{\overline{A}} \cap A = \emptyset$ to compute $\boldsymbol{P}(\widehat{A} \to B \,|\, A) = \dfrac{\boldsymbol{P}((\widehat{\overline{A}} \cup B) \cap A)}{\boldsymbol{P}(A)} = \dfrac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(A)} = \boldsymbol{P}(B \,|\, A) = c$.

c) Finally, $\boldsymbol{P}(\widehat{A} \to B \,|\, \widehat{A}) = \boldsymbol{P}(B \,|\, \widehat{A}) = \widehat{c}$, by (8). Note that $\widehat{c} = c^+$, since $F^- = 0$.

Case 2. (False negatives only)

Assume $r^+ = 0$. We proceed as in Case 1, except that now $\widehat{A} \subseteq A$. Set $\gamma = r^-$, so that $A_\gamma = A_{r^-} = A \setminus F^- = \widehat{A} = A^-$. We compute the same conditional probabilities as in Case 1. Note that $\dot{\cup}$ indicates the union is disjoint.

a) $\boldsymbol{P}(A \to B \,|\, \widehat{A}) = \dfrac{\boldsymbol{P}((\overline{A} \cup B) \cap \widehat{A}}{\boldsymbol{P}(\widehat{A})} = \dfrac{\boldsymbol{P}(\widehat{A} \cap B)}{\boldsymbol{P}(\widehat{A})} = \boldsymbol{P}(B \,|\, \widehat{A}) = \widehat{c}$. Note that $\widehat{c} = c^-$, since $F^+ = 0$.

b) From $\widehat{A} = A^-$ and $X^- \subseteq A$ we get $\boldsymbol{P}(\widehat{A} \to B \,|\, A) = \boldsymbol{P}(A^- \to B \,|\, A) = \dfrac{\boldsymbol{P}((X^- \dot{\cup} \overline{A^-}) \cap A))}{\boldsymbol{P}(A)} = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^-)}{\boldsymbol{P}(A)}$. To use (9) and express this in terms of a generalized implication we need γ to satisfy $\boldsymbol{P}(F^-) = \gamma \boldsymbol{P}(\overline{A^-})$ and $\boldsymbol{P}(A) = \boldsymbol{P}(A^-) + \gamma \boldsymbol{P}(\overline{A^-})$. However, $r^- = \boldsymbol{P}(\overline{\widehat{A}} \,|\, A) = \boldsymbol{P}(\overline{A^-} \,|\, A)$ gives $\boldsymbol{P}(F^-) = r^- \boldsymbol{P}(A)$, so $\gamma = r^-$ does not work. The solution is to use the "Bayesian converse" of r^-, i.e., $\widetilde{r}^- = \boldsymbol{P}(A \,|\, \overline{\widehat{A}})$, as defined above. Then $\boldsymbol{P}(F^-) = \widetilde{r}^- \boldsymbol{P}(\overline{\widehat{A}}) = \widetilde{r}^- \boldsymbol{P}(\overline{A^-})$. Since $F^- \subseteq \overline{\widehat{A}}$, we can define $A_{\widetilde{r}^-} = \widehat{A} \cup F^- = A$. Then from (9) we get $\boldsymbol{P}(\widehat{A} \to B \,|\, A) = \boldsymbol{P}(\widehat{A} \xrightarrow{\widetilde{r}^-} B)$.

c) $\boldsymbol{P}(\widehat{A} \to B \,|\, \widehat{A}) = \widehat{c}$, by (8).

Table 1 collects up the results of Case 1 and Case 2, and shows a "duality" between the cases $r^- = 0$ and $r^+ = 0$.

		Case 1 $r^- = 0$	**Case 2** $r^+ = 0$
–	$\boldsymbol{P}(A \to B \,\|\, A) =$	c	c
a)	$\boldsymbol{P}(A \to B \,\|\, \widehat{A}) =$	$\boldsymbol{P}(A \xrightarrow{r^+} B)$	$\widehat{c}$
b)	$\boldsymbol{P}(\widehat{A} \to B \,\|\, A) =$	c	$\boldsymbol{P}(\widehat{A} \xrightarrow{\widetilde{r}^-} B)$
c)	$\boldsymbol{P}(\widehat{A} \to B \,\|\, \widehat{A}) =$	$\widehat{c}$	$\widehat{c}$

Table 1: Duality between F^+, F^-

Case 3. (Both false positives and false negatives)

If $r^-, r^+ > 0$, then $\widehat{A}$ will be distict from both A^- and A^+, so there are 16 cases of the form $\boldsymbol{P}(W \to B \,|\, Y)$, for $W, Y \in \{A, \widehat{A}, A^+, A^-\}$. Nine of these satisfy $Y \subseteq W$, in which case $\boldsymbol{P}(W \to B \,|\, Y) = \boldsymbol{P}(B \,|\, Y) = z$, where z is the corresponding element of $\{c, \widehat{c}, c^+, c^-\}$. Of these nine, four are immediate consequences of (8). The remaining five are analogous to $\boldsymbol{P}(A^+ \to B \,|\, A^-) = \dfrac{\boldsymbol{P}((X^+ \cup \overline{A^+}) \cap A^-)}{\boldsymbol{P}(A^-)} = \dfrac{\boldsymbol{P}(X^-)}{\boldsymbol{P}(A^-)} = c^-$. The remaining seven cases, where $W \not\subseteq Y$ and $Y \not\subseteq W$, are collected in Table 2. The relevant computations are shown below. As before, $\dot{\cup}$ indicates the union is disjoint.

a) $\boldsymbol{P}(A^- \to B \mid A) = \dfrac{\boldsymbol{P}((X^- \dot\cup \overline{A^-}) \cap A)}{\boldsymbol{P}(A)} = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^-)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^-)} = A^- \xrightarrow{\boldsymbol{\gamma}} B$, for γ satisfying $\boldsymbol{P}(F^-) = \gamma(1 - a^-)$. However, we currently only have $\boldsymbol{P}(F^-)$ $= r^- a = \widetilde{r}^-(1 - \widehat{a})$. The solution is to let $\gamma = u^- = \boldsymbol{P}(F^- \mid \overline{A^-})$, as defined above. Then $u^- = \dfrac{\widetilde{r}^-(1 - \widehat{a})}{1 - a^-} = \dfrac{r^- a}{1 - a^-}$. Then we get $\boldsymbol{P}(\widehat{A} \to B \mid A) =$ $\boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}^-} B)$.

b) $\boldsymbol{P}(\widehat{A} \to B \mid A) = \dfrac{\boldsymbol{P}((\widehat{X} \dot\cup \widehat{\overline{A}}) \cap A)}{\boldsymbol{P}(A)} = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^-)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^-)} = A^- \xrightarrow{\boldsymbol{u}^-} B$.

c) $\boldsymbol{P}(A \to B \mid \widehat{A}) = \dfrac{\boldsymbol{P}((X \dot\cup \overline{A}) \cap \widehat{A})}{\boldsymbol{P}(\widehat{A})} = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^+)}$. Similar to case (a) above, we use $\boldsymbol{P}(F^+) = r^+(1 - a) = \widetilde{r}^+ \widehat{a}$, to define $u^+ = \dfrac{r^+(1 - a)}{1 - a^-} =$ $\dfrac{\widetilde{r}^+ \widehat{a}}{1 - a^-}$. Then $\boldsymbol{P}(A \to B \mid \widehat{A}) = \boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}^+} B)$.

d) $\boldsymbol{P}(A^- \to B \mid \widehat{A}) = \dfrac{\boldsymbol{P}((X^- \dot\cup \overline{A}) \cap \widehat{A})}{\boldsymbol{P}(\widehat{A})} = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^+)} = \boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}^+} B)$, as in the previous case.

e) $\boldsymbol{P}(A \to B \mid A^+) = \dfrac{\boldsymbol{P}((X \dot\cup \overline{A}) \cap A^+)}{\boldsymbol{P}(A^+)} = \dfrac{\boldsymbol{P}(X) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A) + \boldsymbol{P}(F^+)} = \boldsymbol{P}(A \xrightarrow{\boldsymbol{r}^+} B)$, since $\boldsymbol{P}(F^+) = r^+ \boldsymbol{P}(\overline{A})$.

f) $\boldsymbol{P}(\widehat{A} \to B \mid A^+) = \dfrac{\boldsymbol{P}((\widehat{X} \dot\cup \widehat{\overline{A}}) \cap A^+)}{\boldsymbol{P}(A^+)} = \dfrac{\boldsymbol{P}(\widehat{X}) + \boldsymbol{P}(F^-)}{\boldsymbol{P}(\widehat{A}) + \boldsymbol{P}(F^-)} = \boldsymbol{P}(A \xrightarrow{\widetilde{\boldsymbol{r}}^-} B)$, since $\boldsymbol{P}(F^-) = \widetilde{r}^- \boldsymbol{P}(\widehat{\overline{A}})$.

g) $\boldsymbol{P}(A^- \to B \mid A^+) = \dfrac{\boldsymbol{P}((X^- \dot\cup \overline{A^-}) \cap A^+)}{\boldsymbol{P}(A^+)} = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^-) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^-) + \boldsymbol{P}(F^+)}$.

We give three ways to express $\boldsymbol{P}(A^- \to B \mid A^+)$ as a generalized implication, by grouping the terms in this last ratio differently.

First, recall that $\boldsymbol{P}(F^-) + \boldsymbol{P}(F^+) = u^-(1 - a^-) + u^+(1 - a^-)$. So we use $u = u^- + u^+$, as defined above, to get $\boldsymbol{P}(A^- \to B \mid A^+) = \boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}} B)$.

Second, note that $\boldsymbol{P}(A \cap (B \cup F^-)) = \boldsymbol{P}(X^-) + \boldsymbol{P}(F^-)$. Then, using $\boldsymbol{P}(F^+) = r^+ \boldsymbol{P}(\overline{A})$, we get $\boldsymbol{P}(A^- \to B \mid A^+) = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^-) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^-) + \boldsymbol{P}(F^+)}$ $= \dfrac{\boldsymbol{P}(A \cap (B \cup F^-)) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A) + \boldsymbol{P}(F^+)} = \boldsymbol{P}(A \xrightarrow{\boldsymbol{r}^+} (B \cup F^-))$.

Third, note that $\boldsymbol{P}(\widehat{A} \cap (B \cup F^+)) = \boldsymbol{P}(X^-) + \boldsymbol{P}(F^+)$. Then, from $\boldsymbol{P}(F^-)$ $= \widetilde{r}^-(1 - \widehat{a})$, we get that $\boldsymbol{P}(A^- \to B \mid A^+) = \dfrac{\boldsymbol{P}(X^-) + \boldsymbol{P}(F^-) + \boldsymbol{P}(F^+)}{\boldsymbol{P}(A^-) + \boldsymbol{P}(F^-) + \boldsymbol{P}(F^+)}$

$$= \frac{\boldsymbol{P}(\widehat{A} \cap (B \cup F^+)) + \boldsymbol{P}(F^-)}{\boldsymbol{P}(\widehat{A}) + \boldsymbol{P}(F^-)} = \boldsymbol{P}(\widehat{A} \xrightarrow{\widetilde{\boldsymbol{r}}^-} (B \cup F^+)).$$

Finally, we note that all of the above can be expressed in terms of the classical conditionals $\boldsymbol{P}(B \cup Y \,|\, W)$, where $Y \in \{F^-, F^+, F\}$ and $W \in \{A, A^-, \widehat{A}, A^+\}$. Table 2 collects up all the non-trivial cases, i.e., those that do not lead to $\boldsymbol{P}(W \to B \,|\, Y) \in \{c, c^-, \widehat{c}, c^+\}$, for $W, Y \in \{A, A^-, \widehat{A}, A^+\}$.

	Case 3 $r^+, r^- \geq 0$				
a)	$\boldsymbol{P}(A^- \to B \,\|\, A)$	$=$	$\boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}^-} B)$	$=$	$\boldsymbol{P}(B \cup F^- \,\|\, A)$
b)	$\boldsymbol{P}(\widehat{A} \to B \,\|\, A)$	$=$		$=$	$\boldsymbol{P}(B \cup F \,\|\, A)$
c)	$\boldsymbol{P}(A \to B \,\|\, \widehat{A})$	$=$	$\boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}^+} B)$	$=$	$\boldsymbol{P}(B \cup F^+ \,\|\, \widehat{A})$
d)	$\boldsymbol{P}(A^- \to B \,\|\, \widehat{A})$	$=$		$=$	$\boldsymbol{P}(B \cup F \,\|\, \widehat{A})$
e)	$\boldsymbol{P}(A \to B \,\|\, A^+)$	$=$	$\boldsymbol{P}(A \xrightarrow{\boldsymbol{r}^+} B)$	$=$	$\boldsymbol{P}(B \cup F^+ \,\|\, A^+)$
f)	$\boldsymbol{P}(\widehat{A} \to B \,\|\, A^+)$	$=$	$\boldsymbol{P}(\widehat{A} \xrightarrow{\widetilde{\boldsymbol{r}}^-} B)$	$=$	$\boldsymbol{P}(B \cup F^- \,\|\, A^+)$
g)	$\boldsymbol{P}(A^- \to B \,\|\, A^+)$	$=$	$\boldsymbol{P}(A^- \xrightarrow{\boldsymbol{u}} B)$		
		$=$	$\boldsymbol{P}(A \xrightarrow{\boldsymbol{r}^+} (B \cup F^-))$		
		$=$	$\boldsymbol{P}(\widehat{A} \xrightarrow{\widetilde{\boldsymbol{r}}^-} (B \cup F^+))$	$=$	$\boldsymbol{P}(B \cup F \,\|\, A^+)$

Table 2: Symmetries in generalized implications

8.3. Duality between false positives and false negatives

Note that $\widehat{A}$ might contain false positives and/or might be missing false negatives, so in general, $\widehat{A} = (A \cup F^+) \setminus F^-$. Similarly, $A = (\widehat{A} \cup F^-) \setminus F^+$. The relationship between the relevant sets can be shown diagrammatically:

$$\begin{tikzcd} & \boldsymbol{A} = A^- \cup \boldsymbol{F}^- \arrow[dr, phantom, "\subseteq" sloped] & \\ A^- \arrow[ur, phantom, "\subseteq" sloped] \arrow[dr, phantom, "\subseteq" sloped] & & A^+ \\ & \widehat{\boldsymbol{A}} = A^- \cup \boldsymbol{F}^+ \arrow[ur, phantom, "\subseteq" sloped] & \end{tikzcd} \tag{11}$$

The "duality" between $A, \widehat{A}$ that is apparent in (11) leads to the symmetries in the γ-values seen in Table 2. Figure 3 can be viewed as a decoration of (11), and a graphical visualization of the symmetries present in Table 2 (as well as the nine cases excluded from Table 2 with $\gamma = 0$, as mentioned at the beginning of Case 3 above). In Figure 3, note that if $W \supseteq Y$, then we get a classical conditional with $\gamma = 0$ (the **cyan** "arrows"), and if $W \subset Y$ then we get a non-trivial γ-value (the **orange** "arrows", which correspond to some entry in Table 2). Note that the specific γ-values we get are symmetric across the horizontal (as A and $\widehat{A}$ are "dual"). Furthermore, the symmetry amongst the γ-values is exactly the same as the symmetry between $B \cup F^-$, $B \cup F^+$, and $B \cup F$, as can be seen in Table 2.

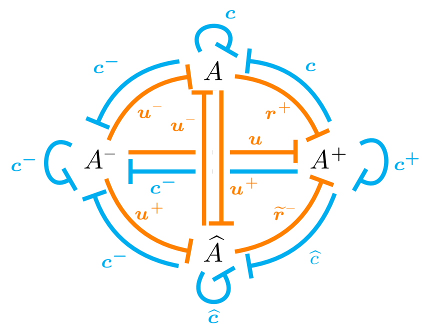


Figure 3: Symmetry for γ-values. The **orange** symbol $W \overset{\gamma}{\dashv} Y$ represents the equality $\boldsymbol{P}(W \to B \mid Y) = \boldsymbol{P}(W \xrightarrow{\gamma} B)$. The **cyan** symbol $W \overset{c}{\dashv} Y$ represents the equality $\boldsymbol{P}(W \to B \mid Y) = \boldsymbol{P}(W \xrightarrow{0} B) = c$.

9. Generalized implications and applications to logic

9.1. Paradoxes of material implication

We will look at a couple simple examples of "paradoxes of material implication" (e.g., see [38]), and show what light is shed on these by replacing $\longrightarrow$ with $\xrightarrow{\gamma}$. The sentence $\bot \to B \equiv \top$ is considered problematic since, intuitively, it should be false. In terms of events, the relevant generalized implication would be $\emptyset \xrightarrow{\gamma} B$. Of course, the Bayesian conditional c is not defined for this case, but $x = 0$ nonetheless. For every $\gamma > 0$, we see from (2) that $g(0, 0, \gamma) = 1$, so no choice of γ will mitigate this "paradox". The same behaviour occurs if we approximate $\bot \xrightarrow{\gamma} B$ with $A \xrightarrow{\gamma} B$ for $0 < a \ll 1$. Then c is defined, and if $\gamma = 0$, then $g(a, x, 0) = c \in [0, 1]$, and if $\gamma = 1$, then $g(a, x, 1) = m \in [1 - a, 1]$, so if $a \approx 0$ then $m \approx 1$. So again, no choice of γ will mitigate this "paradox".

The sentence $A \to \neg A \not\equiv \bot$ is problematic since, intuitively, it should also be false. Clearly, $c = x = 0$ for any choice of $A \neq \emptyset$. Hence, $\gamma = 0$ gives the intuitively correct probability of 0, for all $A \neq \emptyset$. However, $\gamma = 1$ gives the probability of this sentence to be $\boldsymbol{P}(\overline{A})$. Since the "paradox" occurs when $g(a, 0, \gamma) > 0$, the average deviation of g from the intuitively correct value of 0 (over all possible choices of A and γ) can be computed with an integral: $\int_0^1 \int_0^1 g(a, 0, \gamma)\, da\, d\gamma = 2 - \frac{\pi^2}{6} \approx 0.3551$, and $\frac{1}{0.5} \int_0^{0.5} \int_0^1 g(a, 0, \gamma)\, da\, d\gamma \approx 0.2609$.

9.2. Lewis triviality

The thesis that the probability of an indicative conditional $\Rightarrow$, $\boldsymbol{P}(A \Rightarrow B)$, equals the probability of the consequent conditional on its antecedent, $\boldsymbol{P}(B \mid A)$ (as long as $\boldsymbol{P}(A) > 0$), has been attributed to both Adams [1] and Stalnaker

[35], and [23] points out that it was first observed in [33]. There are many arguments to support this thesis in both the philosophical and psychological literature, but some of the most serious arguments against it are the Lewis triviality results [24] (and their many refinements). Lewis gave simple conditions under which we are led to an untenable result, e.g., $\boldsymbol{P}(A \Rightarrow B) = \boldsymbol{P}(B)$. If we try something analogous to Adams' thesis, but with generalized implications, by setting $\boldsymbol{P}(A \xrightarrow{\gamma} B) = \boldsymbol{P}(A \xrightarrow{\boldsymbol{\beta}} B)$, we easily derive from (2) that $(\gamma - \beta)(1-a)(a-x) = 0$, and hence, $\gamma = \beta$. However, if we don't insist on equality with 0, but rather only $(\gamma - \beta)(1-a)(a-x) \leq \epsilon$, then setting $\beta = 0$ gives $\gamma \leq \dfrac{\epsilon}{(1-a)(a-x)}$. This will give the set of γ for which this version of Adams' thesis is satisfied "within ϵ". The Lewis arguments are avoided because we are not imposing the problematic equality $\boldsymbol{P}(A \Rightarrow B) = \boldsymbol{P}(B \mid A)$, but rather looking for deviations from the true equation (8), i.e., solutions to $\boldsymbol{P}(A \Rightarrow B \mid A_\gamma) - \boldsymbol{P}(B \mid A) \leq \epsilon$.

10. Discussion

Our investigation of the properties of the generalized implication revealed several ways for expressing and computing bounds. In particular, it showed the close relationship between probabilistic material implication and Bayesian conditional, especially when used for inference with rules like modus ponens. Paired with the generalized rule, inference principles like modus ponens can generate bounds on the consequent that depend both on the γ-value and the rule bounds, and the theorem about implication chains shows that it is possible to calculate inferences from any mixture of rule bounds and γ-values (recursively in the worst case when clippings happen). This then enables mixed rule-based systems with different γ-values for rules, e.g., high γ-values for rules that are based on little data and are more conceptual in nature, and low γ-values for rules for which sufficient data exists, or for which the Bayesian approach is more appropriate. In general, one possible application of the γ-parameter would be to make it a function of the amount of data from which rules have been extracted: the more data is available, the closer the representation should go to the normatively correct Bayesian end point.

Our investigation into possible meanings for the parameter γ led to an interpretation for both γ, and the probability of the generalized implication it defines, as conditional probabilities. This followed naturally from the observation that the γ-parameter determines how much weight is given to the complement of the rule's antecedent: higher γ-values take more of the complement into account. This can be useful, for example, in cases with false negative rates where we care about the adjustments to be made to the probability of the rule $\boldsymbol{P}(A \to B)$ since we only have the "measured" $\widehat{A}$ (with false negatives). For to be able to make any inferences like *modus ponens* we cannot use $\boldsymbol{P}(A \to B)$ if A is not given but only the measured $\widehat{A}$ version is available. Hence, we must use the implication with the measured antecedent $\boldsymbol{P}(\widehat{A} \to B)$ given the true A, i.e.,

$\boldsymbol{P}(\widehat{A} \to B \mid A)$, which we showed to be the same as $\boldsymbol{P}(\widehat{A} \xrightarrow{\gamma} B)$ for a specific γ, namely $\widetilde{r}^{-}$ (which considers (the size of) those parts of $\widehat{\widehat{A}}$ that contain the false negatives).[1] This just corresponds to the claim that Bayes' is the normatively correct way to capture statistical inference, and we should not actually do MP with material implication.

Another interesting possible interpretation of γ would be to situate the semantics of the generalized implication in the more abstract context afforded by [12], in which a logico-algebraic theory of conditionals was developed, with the symbol $B \mid A$ is taken as primitive. They presented a Boolean algebra of the conditionals $B \mid A$, and defined an order $\leq$ in this Boolean algebra in the standard way. Then they showed the following: $B \mid A \leq (A \to B) \mid \top$ and $B \mid A = (A \to B) \mid A \leq (A \to B) \mid (A \vee A')$. The expressions in these two inequalities correspond exactly to those in (8) and (9) above, and hence, we see that in the context of [12], these generalized implications are all comparable (in this Boolean algebra). However, [12] does not consider these intermediate cases $(0 < \gamma < 1)$ as implication operators.

Future work should proceed in the following ways. First, the connection between γ-values, quantity of data, and measurement errors should be developed further to find practical applications where some γ strictly between 0 and 1 gives an optimal inference. Second, it would be interesting to investigate the proof theory of the generalized implication. A complete proof system for probability logic is put forward in [31]. Inference in conditional probability logic was studied in [27]. An interesting and potentially important next step would be to extend the Gentzen-style proof system of [4] by adding all of the generalized implications $\xrightarrow{\gamma}$, for $0 \leq \gamma \leq 1$, to the logic. The goal would be to have a formal proof system that includes both probabilistic classical propositional logic and Bayesian reasoning as special cases. Finally, it would be interesting to explore the usefulness of the generalized implication in analyzing the psychological data from studies about how humans interpret implications in natural language.

11. Conclusion

The goal of this paper was to investigate the relationship of probabilistic material implications and Bayesian conditionals in an effort to understand better the modeling of "if-then" rules. For this purpose we utilized a generalization, $A \xrightarrow{\gamma} B$, which subsumes both formalisms as special cases, revealing them as end points of a one-parameter family, for $\gamma \in [0, 1]$, in the space of all possible logical rules. We then provided methods for obtaining bounds for different ap-

[1] The method of proportionally considering evidence to the contrary may seems to bear some resemblance to approaches for handling conflicting evidence in the Dempster-Shafer (DS) theory evidence, specifically in the case of Dempster's rule of combination. However, note that possible conflicts with Dempster's rule of combination are the result of conflicts among evidence sources, whereas in the above case we are correcting for faulty, noisy, or imprecise assessments of a single source.

plications of the generalized implication, focussing on *modus ponens* and *modus tollens*, but also on rule converses $B \xrightarrow{\gamma} A$, which showed the potential of using implication and Bayesian conditionals in a uniform manner for various types of inferences, in particular, generalized implication chains. Most importantly, we introduced a novel interpretation of the generalized implication in terms of conditional probabilities, and connected this interpretation to (the probabilities of) false positives and false negatives.

12. Appendix

Proof of Theorem 3:

Proof. We first note that $Y^i_{\boldsymbol{L}} = Y^i(g^i_{\boldsymbol{L}})$, $Y^i_{\boldsymbol{U}} = Y^i(g^i_{\boldsymbol{U}})$, $Z^i_{\boldsymbol{L}} = Z^i(g^i_{\boldsymbol{U}})$, $Z^i_{\boldsymbol{U}} = Z^i(g^i_{\boldsymbol{L}})$. To prove the claim we give a straightforward induction on n.

For the base case $n = 1$, we just need simple substitution and algebra: $b^1_{\boldsymbol{L}} = a_{\boldsymbol{L}} Y^0_{\boldsymbol{L}} - Z^0_{\boldsymbol{U}} = (1-\gamma)a_{\boldsymbol{L}} g^0_{\boldsymbol{L}} + \gamma(a_{\boldsymbol{L}} + g^0_{\boldsymbol{L}} - 1)$, which is just (6). Similarly, $b^1_{\boldsymbol{U}} = 1 - a_{\boldsymbol{L}}(1 - Y^0_{\boldsymbol{U}}) - Z^0_{\boldsymbol{L}} = 1 - a_{\boldsymbol{L}} + (1-\gamma)a_{\boldsymbol{L}} g^0_{\boldsymbol{U}} + \gamma(a_{\boldsymbol{L}} + g^0_{\boldsymbol{U}} - 1)$, which is just (7).

For the inductive step, we get

$$
\begin{aligned}
b^{n+1}_{\boldsymbol{L}} = b^n_{\boldsymbol{L}} Y^n_{\boldsymbol{L}} - Z^n_{\boldsymbol{U}} &= \left[a_{\boldsymbol{L}} \textstyle\prod_{i=0}^{n-1} Y^i_{\boldsymbol{L}} - \sum_{i=0}^{n-1} \left(Z^i_{\boldsymbol{U}} \prod_{j=i+1}^{n-1} Y^j_{\boldsymbol{L}} \right) \right] Y^n_{\boldsymbol{L}} - Z^n_{\boldsymbol{U}} \\
&= a_{\boldsymbol{L}} \textstyle\prod_{i=0}^{n} Y^i_{\boldsymbol{L}} - \sum_{i=0}^{n} \left(Z^i_{\boldsymbol{U}} \prod_{j=i+1}^{n-1} Y^j_{\boldsymbol{L}} \right) Y^n_{\boldsymbol{L}} - Z^n_{\boldsymbol{U}} \\
&= a_{\boldsymbol{L}} \textstyle\prod_{i=0}^{n} Y^i_{\boldsymbol{L}} - \sum_{i=0}^{n} \left(Z^i_{\boldsymbol{U}} \prod_{j=i+1}^{n} Y^j_{\boldsymbol{L}} \right) Y^n_{\boldsymbol{L}},
\end{aligned}
$$

as desired, and similarly for $b^{N+1}_{\boldsymbol{U}}$.

Recall that $b^i_{\boldsymbol{L}} = \cdots$, $b^i_{\boldsymbol{U}} = \cdots$ are automatically clipped to remain in $[0,1]$, as per Definition 2, but that this clipping is lost if we combine and rearrange the RHS expressions for these quantities, rather than referring to the LHS quantities themselves. So the above formulas assume no clipping was necessary at intermediate stages of the recursion. $\square$

Proof of Lemma 6:

Proof. We just use equation (3) (with g, γ interchanged with $\hat{g}, \hat{\gamma}$) to write all of the γ^i, g^i pairs in terms of a single fixed $\hat{\gamma}$ (of our choice), and the corresponding (newly computed) $\hat{g}^i$. Then apply Theorem 3. $\square$

Proof of Lemma 7:

Proof. Note that everything is specified except for the bounds on g. Since γ, $a_{\boldsymbol{L}}$, $b_{\boldsymbol{L}} = b^N_{\boldsymbol{L}}$, and $b_{\boldsymbol{U}} = b^N_{\boldsymbol{U}}$ are all fixed, equations (6), (7) give two linear equations each with one unknown. $\square$

Proof of Lemma 8:

Proof. Just combine the ideas in the proofs of Lemmas 6 and 7. $\square$

References

[1] Ernest Adams. The logic of conditionals. *Inquiry: An Interdisciplinary Journal of Philosophy*, 8(1-4):166–197, 1965. doi: 10.1080/00201746508601430.

[2] George Boole. *George Boole's Collected Logical Works*. Open Court, 1916.

[3] Marija Boričić. Suppes-style sequent calculus for probability logic. *Journal of Logic and Computation*, 27(4):1157–1168, 10 2015. ISSN 0955-792X. doi: 10.1093/logcom/exv068. URL https://doi.org/10.1093/logcom/exv068.

[4] Marija Boričić. Sequent calculus for classical logic probabilized. *Archive for Mathematical Logic*, 58(1/2):119 – 136, 2019. ISSN 09335846.

[5] Ruth M.J. Byrne and P.N. Johnson-Laird. 'if' and the problems of conditional reasoning. *Trends in Cognitive Sciences*, 13(7):282–287, 2009. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2009.04.003. URL https://www.sciencedirect.com/science/article/pii/S1364661309001144.

[6] Philip Calabrese. An algebraic synthesis of the foundations of logic and probability. *Information Sciences*, 42(3):187–237, 1987. ISSN 0020-0255. doi: https://doi.org/10.1016/0020-0255(87)90023-5. URL https://www.sciencedirect.com/science/article/pii/0020025587900235.

[7] Ivano Ciardelli and Adrian Ommundsen. Probabilities of conditionals: updating adams. *Noûs*, 11 2022. doi: https://doi.org/10.1111/nous.12437. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/nous.12437.

[8] Igor Douven and Richard Dietz. A puzzle about Stalnaker's hypothesis. *Topoi*, 30(1):31–37, 2011. doi: 10.1007/s11245-010-9082-3.

[9] Didier Dubois and Henri Prade. The logical view of conditioning and its application to possibility and evidence theories. *International Journal of Approximate Reasoning*, 4(1):23–46, 1990. ISSN 0888-613X. doi: https://doi.org/10.1016/0888-613X(90)90007-O. URL https://www.sciencedirect.com/science/article/pii/0888613X9090007O.

[10] Jonathan Evans and David Over. *If*. Oxford University Press, 10 2004. ISBN 9780198525134. doi: 10.1093/acprof:oso/9780198525134.001.0001. URL https://doi.org/10.1093/acprof:oso/9780198525134.001.0001.

[11] Ronald Fagin and Joseph Y. Halpern. Uncertainty, belief, and probability. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'89, page 1161–1167, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[12] Tommaso Flaminio, Lluis Godo, and Hykel Hosni. Boolean algebras of conditionals, probability and logic. *Artificial Intelligence*, 286: 103347, 2020. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2020.103347. URL https://www.sciencedirect.com/science/article/pii/S000437022030103X.

[13] Alan M. Frisch and Peter Haddawy. Anytime deduction for probabilistic logic. *Artificial Intelligence*, 69(1):93–122, 1994. ISSN 0004-3702. doi: https://doi.org/10.1016/0004-3702(94)90079-5. URL https://www.sciencedirect.com/science/article/pii/0004370294900795.

[14] Nataniel Greene. An overview of conditionals and biconditionals in probability. In *MATH'08: Proceedings of the American Conference on Applied Mathematics*, page 171–177, 2008.

[15] Rolf Haenni, Jan-Willem Romeijn, Gregory Wheeler, and Jon Williamson. *Probabilistic Logics and Probabilistic Networks*. Dordrecht, Netherland: Synthese Library, 01 2010. ISBN 978-94-007-0007-9. doi: 10.1007/978-94-007-0008-6.

[16] Theodore Hailperin. Best possible inequalities for the probability of a logical function of events. *The American Mathematical Monthly*, 72(4):343–359, 1965. ISSN 00029890, 19300972. URL http://www.jstor.org/stable/2313491.

[17] Theodore Hailperin. Probability logic. *Notre Dame Journal of Formal Logic*, 25(3):198 – 212, 1984. doi: 10.1305/ndjfl/1093870625. URL https://doi.org/10.1305/ndjfl/1093870625.

[18] Theodore Hailperin. *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications*. Lehigh University Press, 1996.

[19] Joseph Y. Halpern. *Reasoning about Uncertainty*. The MIT Press, 2017. ISBN 0262533804.

[20] Michael Jahn and Matthias Scheutz. Investigating a generalization of probabilistic material implication and bayesian conditionals. In *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence*, 2023.

[21] Marija Boričić Joksimović. On basic probability logic inequalities. *Mathematics*, 9(12), 2021. ISSN 2227-7390. doi: 10.3390/math9121409. URL https://www.mdpi.com/2227-7390/9/12/1409.

[22] Justin Khoo. *The Meaning of If*. Oxford University Press, 06 2022. ISBN 9780190096700. doi: 10.1093/oso/9780190096700.001.0001. URL https://doi.org/10.1093/oso/9780190096700.001.0001.

[23] Justin Khoo and Matthew Mandelkern. Triviality Results and the Relationship between Logical and Natural Languages. *Mind*, 128(510):485–526, 05 2018. ISSN 0026-4423. doi: 10.1093/mind/fzy006. URL https://doi.org/10.1093/mind/fzy006.

[24] David Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315, 1976. ISSN 00318108, 15581470. URL http://www.jstor.org/stable/2184045.

[25] David Lewis. Probabilities of conditionals and conditional probabilities ii. *Philosophical Review*, 95(4):581–589, 1986. doi: 10.2307/2185051.

[26] Hung Nguyen, Masao Mukaidono, and Vladik Kreinovich. Probability of implication, logical version of bayes theorem, and fuzzy logic operations. In *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291)*, volume 1, pages 530 – 535, 02 2002. ISBN 0-7803-7280-8. doi: 10.1109/FUZZ.2002.1005046.

[27] Gernot D. Kleiter Niki Pfeifer. Inference in conditional probability logic. *Kybernetika*, 42(4):391–404, 2006. URL http://eudml.org/doc/33813.

[28] Nils J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28(1):71–87, 1986. ISSN 0004-3702. doi: https://doi.org/10.1016/0004-3702(86)90031-7. URL https://www.sciencedirect.com/science/article/pii/0004370286900317.

[29] Nils J. Nilsson. Probabilistic logic revisited. *Artificial Intelligence*, 59(1):39–42, 1993. ISSN 0004-3702. doi: https://doi.org/10.1016/0004-3702(93)90167-A. URL https://www.sciencedirect.com/science/article/pii/000437029390167A.

[30] Zoran Ognjanovi, Miodrag Rakovi, and Zoran Markovi. *Probability Logics: Probability-Based Formalization of Uncertain Reasoning*. Springer Publishing Company, Incorporated, 1st edition, 2016. ISBN 3319470116.

[31] J. Paris and A. Vencovská. Proof systems for probabilistic uncertain reasoning. *The Journal of Symbolic Logic*, 63(3):1007–1039, 1998. ISSN 00224812. URL http://www.jstor.org/stable/2586724.

[32] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann series in representation and reasoning. Elsevier Science, 1988. ISBN 9781558604797. URL https://books.google.com/books?id=AvNID7LyMusC.

[33] Frank Plumpton Ramsey. *The Foundations of Mathematics and Other Logical Essays. Edited by R.B. Braithwaite, with a Pref. By G.E. Moore.* –. London, England: Routledge and Kegan Paul, 1931.

[34] Geza Schay. An algebra of conditional events. *Journal of Mathematical Analysis and Applications*, 24(2):334–344, 1968. ISSN 0022-247X. doi: https://doi.org/10.1016/0022-247X(68)90035-8. URL https://www.sciencedirect.com/science/article/pii/0022247X68900358.

[35] Robert Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, pages 98–112. Oxford: Blackwell, 1968.

[36] K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science*. A Bradford Book. MIT Press, 2012. ISBN 9780262293532. URL https://books.google.com/books?id=c3s4Wa0gBLkC.

[37] Crupi Vincenzo and Iacona Andrea. Three ways of being non-material. *Studia Logica*, 110, 01 2019. doi: 10.1007/s11225-021-09949-y.

[38] Kai von Fintel. Conditionals. In Klaus von Heusinger, Claudia Maienborn, and Paul Portner, editors, *Semantics: An international handbook of meaning*, volume 2, pages 1515—1538. de Gruyter Mouton, Berlin/Boston, 2011.

[39] Carl G. Wagner. Modus tollens probabilized. *The British Journal for the Philosophy of Science*, 55(4):747–753, 2004. ISSN 00070882, 14643537. URL http://www.jstor.org/stable/3541627.