Investigating a Generalization of Probabilistic Material Implication and Bayesian Conditionals

Abstract

Probabilistic "if A then B" rules are typically formalized as Bayesian conditionals $\boldsymbol{P}(B \mid A)$, as many (e.g., Pearl) have argued that Bayesian conditionals are the correct way to think about such rules. However, 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 rule, in particular, for inference chains in graphs.

1 INTRODUCTION

Natural language "if-then" rules cover a wide range of logically distinct conditions such as *analytical truths* ("if x is an even number, x is divisible by two"), *conceptual relationships* ("if x is a human, then x is mortal"), *inductive inferences* ("if the sprinkler is on, the grass will become wet"), *abductive inferences* ("if the light switch is on but the light is off, the light bulb is broken"), *normative constraints* ("if the traffic light is red, you are not allowed to drive"), *hypotheticals* ("if I were to run fast, I would be out of breath"), and *counterfactuals* ("if the driver had been able to brake in time, they would not have killed the deer") (see also Ciardelli and Ommundsen, Crupi and Iacona [2021], Dubois and Prade [1990], Nilsson [1993], Greene [2008]).

Treating conditionals as material implications, however, has long been seen to be problematic [e.g., Adams, 1965] and the widely accepted solution is to view such rules as *probabilistic* and best modeled by the Bayesian conditional $\boldsymbol{P}(B \mid A)$ [e.g., Pearl, 1988]. At the same time, Bayesian conditionals are not appropriate for handling (indicative) conditionals [e.g., see Khoo and Mandelkern, 2018, for a discussion], and cannot be used for inference (e.g., using *modus tollens*) if the probability of the prior is zero: $P(A) = 0$.

Fortunately, there are ways to combine probabilistic logical and Bayesian inference in a way that views probabilistic material implication and Bayesian conditionals as two limit points of a "rule-continuum", $A \xrightarrow{\gamma} B$, expressed by a real-valued parameter $\gamma \in [0, 1]$ (with $\gamma = 1$ for material implication $A \to B$, and $\gamma = 0$ for the Bayesian conditional $A|B$).

The goal of this paper is to investigate the formal properties of the "generalized rule" $\xrightarrow{\gamma}$ and to develop a framework that can be used for reasoning with it. We start with a brief motivation for the generalization and introduce its mathematical form. Then we start to investigate its various properties: functional relationships among various probabilites comparing material implication, Bayesian conditional, and generalize rule, as well as their bounds. Then we consider implication chains and prove a boundary results for generalized implication chains. We conclude with a brief discussion and summary of our findings.

2 MOTIVATION AND RELATED WORK

Efforts to combine probability and logic go back at least to Leibniz, Jakob Bernoulli, and Boole (e.g., see Hailperin [1996], and also Hailperin [1984] 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 $A \to B \equiv \neg A \wedge B$ we get the standard

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

Submitted to the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023). **To be used for reviewing only**.

It was recognized at least as early as Boole [1916] (see also Hailperin [1984]) that bounds are needed when dealing with probabilities of events (i.e., sets representing propositional sentences). Using linear programming, the existence of lower and upper bound functions for the probability of any propositional logic formula was established in Hailperin [1965]. In Hailperin [1984], these results were extended, and probability bounds on *modus ponens* and *hypothetical syllogism* were given.

For our purposes, the most salient motivation for working with imprecise, i.e., bounded, probabilities is that they are *unavoidable* if we want to use *modus ponens* with probabilistic implication rules: given $a = \boldsymbol{P}(A)$ and $m = \boldsymbol{P}(A \to B)$, applying *modus ponens* to infer $b = \boldsymbol{P}(B)$ only yields bounds on $b \in [x, m]$ where $x = \boldsymbol{P}(A \cap B)$. A formal semantic method to compute the bounds on the probability of a given sentence in the predicate calculus was developed in Nilsson [1986] and extended in Fagin and Halpern [1989]. Note that we can also use the Bayesian conditional $B \mid A$ for inference (e.g., instead of material implication) to get bounds on B which were initially presented in Wagner [2004] (see below for details).

An abstract notion of probabilistic logical implication $A \xrightarrow{\gamma} B$ was introduced in Nguyen et al. [2002], and the probability of such a generalized implication was computed as a function of a parameter γ, $0 \leq \gamma \leq 1$:

$$\begin{aligned} \boldsymbol{P}(A \xrightarrow{\gamma} B) &= 1 - \frac{\boldsymbol{P}(A)(1 - \boldsymbol{P}(B \mid A))}{\gamma + (1-\gamma)\boldsymbol{P}(A)} \\ &= 1 - \frac{a(1-c)}{\gamma + (1-\gamma)a} \end{aligned} \tag{2}$$

Alternatively, we can write this generalize rule as

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

which more clearly shows the overall form of the Bayesian conditional with the added probability mass of $\overline{A}$ scaled by γ in proportion in the numerator and denominator.

It is immediately apparent that $\gamma = 0$ corresponds to the Bayesian conditional, and $\gamma = 1$ corresponds to 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.

For the rest of the paper, we will use the following notation and conventions:

Definition 1 (Notation for basic probability variables)**.** Fix A, B, γ and let $X = A \cap B$. Then $a = \boldsymbol{P}(A)$, $b = \boldsymbol{P}(B)$, $x = \boldsymbol{P}(X)$, $m = \boldsymbol{P}(A \to B)$ for *m*aterial implication, $c = \boldsymbol{P}(B \mid A)$ for Bayesian *c*onditional, $g = \boldsymbol{P}(A \xrightarrow{\gamma} B)$ 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. But typically, we work under the assumption that $a \neq 0$, so that c is defined. Furthermore, since a, b, x, m, c are probabilities, we obviously must have $0 \leq a, b, x, m, c, g \leq 1$.

3 THE GENERALIZED IMPLICATION

We start by investigating the relationships among the generalized implication $A \xrightarrow{\gamma} B$, the material implication $A \to B$, the Bayesian conditional $B \mid A$, the premise A and the intersection $A \cap B$.

3.1 FUNCTIONAL RELATIONSHIPS AMONG m, c, a, x, g

We first express all of our basic probabilities in terms of the classical probabilities m, c, a, x:

Lemma 2. *For m, c, a as in Definition 1, we have the following functional relationships:*

i) $m = 1 - a + ac = 1 - a(1-c)$,

ii) $c = \dfrac{m + a - 1}{a} = 1 - \dfrac{1-m}{a}$, *for* $a \neq 0$,

iii) $a = \dfrac{1-m}{1-c}$, *for* $c \neq 1$.

Furthermore, using these we can express $x = \boldsymbol{P}(A \cap B)$ in the following ways:

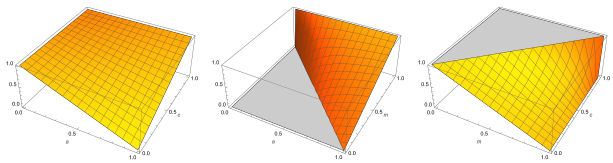
iv) $x = ac = a + m - 1 = \left(\dfrac{1-m}{1-c}\right) c$.

Proof. These are all just restatements of (1), or the definition of c. □

The following relationship between $m = \boldsymbol{P}(A \to B)$ and $c = \boldsymbol{P}(B \mid A)$ is an easy consequence of Lemma 2 and the definition of conditional probability.

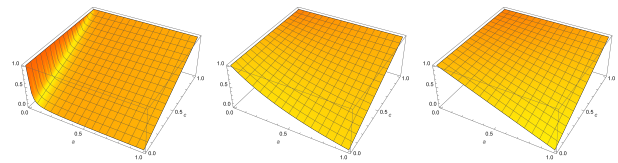
Lemma 3. $\boldsymbol{P}(A \to B) = m \geq c = \boldsymbol{P}(B \mid A)$, *in fact,*

$$m \begin{cases} = 1 & \text{if } a = 0, \\ = c = b & \text{if } a = 1, \\ = c = 1 & \text{if } 0 < a < 1 \text{ and } a = x, \\ > c & \text{if } 0 < a < 1 \text{ and } a \neq x. \end{cases}$$



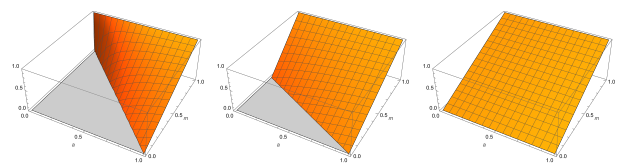
$m(a,c)$ $c(a,m)$ $a(m,c)$

Figure 1: Surfaces coming from m, c, a



$\gamma = 0.01$ $\gamma = 0.5$ $\gamma = 1.0$

Figure 2: $g(\gamma, a, c)$ for various γ



$\gamma = 0.01$ $\gamma = 0.5$ $\gamma = 1.0$

Figure 3: $g(\gamma, a, m)$ for various γ

Proof. For the inequality $m > c$, first note that $a \neq x \Rightarrow c < 1$. Then compute $m - c = (1-a)(1-c) > 0$. □

Note that each of the functions in Lemma 2 gives a surface, but solving $1 - m - a + ac = 0$ for different variables leads to different graphs. These are shown in Figure 1.

The next lemma extends (and subsumes) Lemma 2 by expressing everything in terms of arbitrary γ, g rather than $\gamma = 0, g = c$ and $\gamma = 1, g = m$, as before.

Lemma 4. *Let m, c, a, x, γ, g be as in Definition 1 and Lemma 2. Then*

i) $$g = \frac{ac + \gamma(1-a)}{a + \gamma(1-a)} = \frac{(a+m-1) + \gamma(1-a)}{(1-\gamma)a + \gamma \cdot 1} = \frac{(1-\gamma)x + \gamma m}{(1-\gamma)a + \gamma \cdot 1}$$

ii) $x = (1-\gamma)ag + \gamma(a+g-1)$,

iii) $m = (1-\gamma)ag + \gamma(a+g-1) + (1-a)$,

iv) $c = \dfrac{(1-\gamma)ag + \gamma(a+g-1)}{a}$,

v) $a = \dfrac{\gamma(1-g)}{\gamma(1-g) + g - c}$,

vi) $\gamma = \dfrac{a(g-c)}{(1-a)(1-g)}$,

and for arbitrary $\hat{\gamma}$, with $\hat{g} = g(\hat{\gamma}, x, a)$ from part (i) above,

vii) $$\hat{g} = \frac{[(1-\gamma)ag + \gamma(a+g-1)] + \hat{\gamma}(1-a)}{a + \hat{\gamma}(1-a)},$$

assuming the variables are in the domains of the relevant formula.

Proof. The first six come from (2) and straightforward arithmetic. To get Lemma 4.vii just substitute $x(\gamma, g, a)$ from Lemma 4.ii into $\hat{g} = \hat{g}(\hat{\gamma}, a, x)$ from Lemma 4.ii. □

Note that all the expressions in Lemma 4.i have the form: an interpolation $x \mapsto m$ over an interpolation $a \mapsto 1$. It is in this sense that g interpolates from c to m as γ goes from 0 to 1. It is interesting to note that Lemma 4.ii, the expression for $x(\gamma, g, a)$ is an interpolation from one *expression* of x to another, i.e., $x(a, c) \mapsto x(a, m)$, as γ goes from 0 to 1. The analogous claim is also true for Lemma 4.iii, 4.iv. The point of Lemma 4.vii is that the g defined in terms of one γ can also be defined in terms of another $\hat{\gamma}$ and its corresponding $\hat{g}$. This is just the general version of Lemma 4.i, 4.iii and 4.iv, and will be important later when we look at implication chains.

As above, all of these functions give a surface when graphed. An important point, however, is that all these expressions in Lemma 4.i give *different* functions when considered as functions of the same two variables. (Of course, the same is true for any of the other quantities when expressed in terms of different variables.) For example, Figure 2 shows the graphs $g(\gamma, a, c)$, for various γ, and Figure 3 show the same series for $g(\gamma, a, m)$.

3.2 BOUNDS ON PROBABILITIES

Henceforth, all probabilities will 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 $p_0 \leq p \leq p_1$. 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 5 (The Lower- and Upper-bound operators)**.** Let p_0, p_1 be values for the lower and upper bounds on probability p, i.e., $p_0 \leq p \leq p_1$, but that p_0, p_1 are computed from functions that do not necessarily give values in $[0, 1]$. Then

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

For any probability p, we will write $p \in [p_{\boldsymbol{L}},\ p_{\boldsymbol{U}}]$, and we will denote the width of any probability interval $[p_{\boldsymbol{L}},\ p_{\boldsymbol{U}}]$ as $\Delta p = p_{\boldsymbol{U}} - p_{\boldsymbol{L}}$.

We next compute bounds on one of a, c, m, b given bounds on the other variables by observing that the relevant function is monotone increasing or decreasing (obvious from the graphs above), and then computing the relevant endpoints to get the extreme values:

Lemma 6. *Every function in Lemma 2 and Lemma 4.i-4.vi is of the form $u(v^1, \ldots, v^k)$, for some $u, v^1, \ldots, v^k \in \{a, m, c, x, \gamma, g\}$. Fix such a $u(v^1, \ldots, v^k)$, and fix $v \in \{v^1, \ldots, v^k\}$. Then $\frac{\partial u}{\partial v}(v^1, \ldots, v^k)$ is either always positive, for all $v^1, \ldots, v^k \in (0, 1)$, or it is always negative, for all $v^1, \ldots, v^k \in (0, 1)$, assuming of course, that the partial is defined for these $v^1, \ldots, v^k$.*

Some specific examples of the computations in Lemma 6 are: $\partial m/\partial c = a > 0$, $\partial c/\partial a = (1-m)/a^2 > 0$, $\partial a/\partial m = -1/(1-c) < 0$. We remark that the negative derivatives are to be expected, since they are a consequence of the cyclic chain rule. For example, $\frac{\partial m}{\partial c}\frac{\partial c}{\partial a}\frac{\partial a}{\partial m} = -1$.

We note that $g(\hat{\gamma}, \hat{g}, \gamma, a)$ from Lemma 4.vii satisfies $\frac{\partial g}{\partial a} = \frac{\partial}{\partial a}\frac{[(1-\hat{\gamma})a\hat{g} + \hat{\gamma}(a + \hat{g} - 1)] + \gamma(1-a)}{a + \gamma(1-a)} = \frac{(1-\hat{g})(\hat{\gamma} - \gamma)}{a + \gamma(1-a)}$, so the function $\lambda a\, g(\hat{\gamma}, \hat{g}, \gamma, a)$ is decreasing if $\hat{\gamma} < \gamma$ and increasing if $\hat{\gamma} > \gamma$.

Lemma 7. *For every $a, m, c, x, \gamma, g \in (0, 1)$,*

i) The partial derivatives of m, c, a are either always positive at every point where they are defined, or always negative. In particular,

$$\frac{\partial m}{\partial c} = a, \frac{\partial a}{\partial c} = \frac{1-m}{(1-c)^2}, \frac{\partial c}{\partial m} = \frac{1}{a}, \frac{\partial c}{\partial a} = \frac{1-m}{a^2} > 0,$$
$$\frac{\partial m}{\partial a} = -(1-c), \quad \frac{\partial a}{\partial m} = -\frac{1}{1-c} < 0.$$

ii) Similarly, $\partial x/\partial z$, $\partial g/\partial z$, for $z \in \{m, c, a, x, \gamma\}$, are all either strictly positive at every point in their domain, or strictly negative at every point in their domain, with one exception coming from Lemma 4.vii:

$$\begin{aligned}&\frac{\partial g(\hat{\gamma}, \hat{g}, \gamma, a)}{\partial a}\\ &= \frac{\partial}{\partial a}\frac{[(1-\hat{\gamma})a\hat{g} + \hat{\gamma}(a + \hat{g} - 1)] + \gamma(1-a)}{a + \gamma(1-a)}\\ &= \frac{(1-\hat{g})(\hat{\gamma} - \gamma)}{a + \gamma(1-a)},\end{aligned}$$

which is < 0 if $\hat{\gamma} < \gamma$, $= 0$ if $\hat{\gamma} = \gamma$, > 0 if $\hat{\gamma} > \gamma$.

iii) Hence, all these functions achieve their max and min on $[0, 1]$ are monotonic on $(0, 1)$ in each of their variables, with the one exception that has a min when $\hat{\gamma} = \gamma$.

An easy application of Lemma 6 is the observation that the probabilities of generalized rules are monotonic in γ:

Lemma 8. *If $a, c < 1$, then $\gamma_1 < \gamma_2$ implies $\boldsymbol{P}(A \xrightarrow{\gamma_1} B) < \boldsymbol{P}(A \xrightarrow{\gamma_2} B)$.*

Proof. From $\frac{\partial g(\gamma, a, c)}{\partial \gamma} = \frac{a(1-a)(1-c)}{(a + \gamma(1-a))^2} > 0.$ □

3.3 BOUNDS ON m, c, a, g FROM PARTIAL DERIVATIVES

An immediate consequence of 6 is that to compute the bounds of any of our basic probabilities, in terms of the bounds of other variables, we just need to evaluate at the appropriate endpoints, i.e., the bounds of the independent variables.

Lemma 9. *Fix any function from Lemma 6, $u(v^1, \ldots, v^k)$, with $u, v^1, \ldots, v^k \in \{a, m, c, x, \gamma, g\}$, and $v^i \in [v^i_{\boldsymbol{L}}, v^i_{\boldsymbol{U}}]$. Let $v^i_{\boldsymbol{E}} = \begin{cases} v^i_{\boldsymbol{L}} & \text{if } \partial u/\partial v_i > 0 \\ v^i_{\boldsymbol{U}} & \text{if } \partial u/\partial v_i < 0 \end{cases}$ and $v^i_{\boldsymbol{F}} = \begin{cases} v^i_{\boldsymbol{L}} & \text{if } \partial u/\partial v_i < 0 \\ v^i_{\boldsymbol{U}} & \text{if } \partial u/\partial v_i > 0, \end{cases}$ and note that $\{v^i_{\boldsymbol{E}}, v^i_{\boldsymbol{F}}\} = \{v^i_{\boldsymbol{L}}, v^i_{\boldsymbol{U}}\}$. Then we have $u_{\boldsymbol{L}} = u(v^1_{\boldsymbol{E}}, \ldots, v^k_{\boldsymbol{E}})$ and $u_{\boldsymbol{U}} = u(v^1_{\boldsymbol{F}}, \ldots, v^k_{\boldsymbol{F}})$.*

Lemma 10. *For m, c, a, x, g as in Lemma 2, and any $b = \boldsymbol{P}(B)$ such that $x = \boldsymbol{P}(A \cap B)$,*

i) $m_{\boldsymbol{L}} = m(a_{\boldsymbol{U}}, c_{\boldsymbol{L}}) = 1 - a_{\boldsymbol{U}}(1 - c_{\boldsymbol{L}})$
$m_{\boldsymbol{U}} = m(a_{\boldsymbol{L}}, c_{\boldsymbol{U}}) = 1 - a_{\boldsymbol{L}}(1 - c_{\boldsymbol{U}})$

ii) $c_{\boldsymbol{L}} = c(m_{\boldsymbol{L}}, a_{\boldsymbol{L}}) = (m_{\boldsymbol{L}} + a_{\boldsymbol{L}} - 1)/a_{\boldsymbol{L}}$
$c_{\boldsymbol{U}} = c(m_{\boldsymbol{U}}, a_{\boldsymbol{U}}) = (m_{\boldsymbol{U}} + a_{\boldsymbol{U}} - 1)/a_{\boldsymbol{U}}$

iii) $a_{\boldsymbol{L}} = a(a_{\boldsymbol{U}}, c_{\boldsymbol{L}}) = (1 - m_{\boldsymbol{U}})/(1 - c_{\boldsymbol{L}})$
$a_{\boldsymbol{U}} = a(m_{\boldsymbol{L}}, c_{\boldsymbol{U}}) = (1 - m_{\boldsymbol{L}})/(1 - c_{\boldsymbol{U}})$

And, of course, the bounds on x, g can be computed in the same way, eg.,

iv) $g_{\boldsymbol{L}}(\gamma, a, c) = g(\gamma_{\boldsymbol{L}}, a_{\boldsymbol{U}}, c_{\boldsymbol{L}}) = \frac{a_{\boldsymbol{U}}c_{\boldsymbol{L}} + \gamma_{\boldsymbol{L}}(1 - a_{\boldsymbol{U}})}{a_{\boldsymbol{U}} + \gamma_{\boldsymbol{L}}(1 - a_{\boldsymbol{U}})}$.

An easy example of Lemma 9 in action would be to use $\partial m(a, c)/\partial a < 0$ and $\partial m(a, c)/\partial c > 0$ to get $m_{\boldsymbol{L}} = m(a_{\boldsymbol{U}}, c_{\boldsymbol{L}}) = 1 - a_{\boldsymbol{U}}(1 - c_{\boldsymbol{L}})$, and $m_{\boldsymbol{U}} = m(a_{\boldsymbol{L}}, c_{\boldsymbol{U}}) = 1 - a_{\boldsymbol{L}}(1 - c_{\boldsymbol{U}})$. Another would be $g_{\boldsymbol{L}}(\gamma, a, c) = g(\gamma_{\boldsymbol{L}}, a_{\boldsymbol{U}}, c_{\boldsymbol{L}}) = \frac{a_{\boldsymbol{U}}c_{\boldsymbol{L}} + \gamma_{\boldsymbol{L}}(1 - a_{\boldsymbol{U}})}{a_{\boldsymbol{U}} + \gamma_{\boldsymbol{L}}(1 - a_{\boldsymbol{U}})}$, using $\partial g/\partial \gamma > 0$, $\partial g/a < 0$, $\partial g/c > 0$.

The bounds listed in Lemma 9 can be interpreted geometrically. The graph of $m(a, c)$ is given in Figure 4, and each point on this surface gives a consistent assignment of values to m, c, a. Bounds on a, c correspond to vertical planes, and bounds on m to horizontal planes. It is apparent that the m-bounds can make the set $\{a, c :$ there is a consistent m for $a, c\}$ be non-convex.

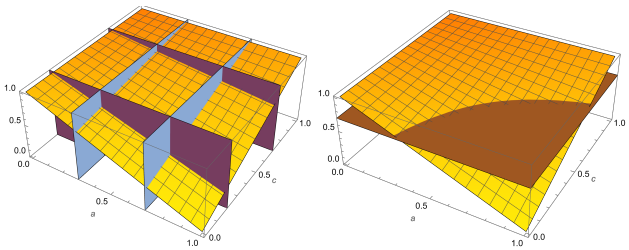


Figure 4: Geometric interpretation of bounds for $m(a,c)$ (a,c-bounds left, m-bounds right).

3.4 BOUNDS ON m, c, g FOR FIXED a, b

For fixed $[a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, $[b_{\boldsymbol{L}},\ b_{\boldsymbol{U}}]$, we want to find $[c_{\boldsymbol{L}},\ c_{\boldsymbol{U}}]$, $[m_{\boldsymbol{L}},\ m_{\boldsymbol{U}}]$, and $[g_{\boldsymbol{L}},\ g_{\boldsymbol{U}}]$. As a first special case, assume $a = a_{\boldsymbol{L}} = a_{\boldsymbol{U}}$, $b = b_{\boldsymbol{L}} = b_{\boldsymbol{U}}$. Then note that $\boldsymbol{P}(A \cap B) = x = ac$, but we do not have c. Note also that x is *not* a function of a, b, but has the following bounds:

$$x_{\boldsymbol{L}} = a + b - 1\,, \qquad x_{\boldsymbol{U}} = \min\{a, b\}. \tag{3}$$

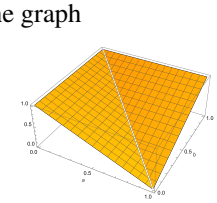
In general, if there are nontrivial probability intervals for a, b, then

$$x_{\boldsymbol{L}} = a_{\boldsymbol{L}} + b_{\boldsymbol{L}} - 1\,, \qquad x_{\boldsymbol{U}} = \min\{a_{\boldsymbol{U}},\ b_{\boldsymbol{U}}\}. \tag{4}$$

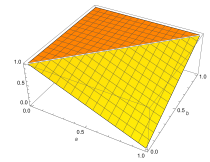
From this we want to compute bounds on c, m.

Of course, we do not have a function $c(a,b)$ to minimize/maximize, but $\frac{\partial}{\partial a}\left(\frac{a+b-1}{a}\right) > 0$, $\frac{\partial}{\partial b}\left(\frac{a+b-1}{a}\right) \geq 0$, and $c_{\boldsymbol{L}}$, $c_{\boldsymbol{U}}$ will occur at the bounds of x, so $[c_{\boldsymbol{L}},\ c_{\boldsymbol{U}}] = [\max\{(a_{\boldsymbol{L}} + b_{\boldsymbol{L}} - 1)/a_{\boldsymbol{L}},\ 0\},\ \min\{b_{\boldsymbol{U}}/a_{\boldsymbol{L}},\ 1\}]$. From this c we can get bounds on $m(a, c(a,b))$ as before, but there might be clipping when recomputed using bounds gotten from $m(a,b)$ as below. Then $[m_{\boldsymbol{L}},\ m_{\boldsymbol{U}}] = [1 - a_{\boldsymbol{U}}(1 - c_{\boldsymbol{L}}),\ 1 - a_{\boldsymbol{L}}(1 - c_{\boldsymbol{U}})]$.

Using $m = 1 - a + x$, x depends on a, b (not a, c), so $[m_{\boldsymbol{L}},\ m_{\boldsymbol{U}}] = [(1 - a + x)_{\boldsymbol{L}},\ (1 - a + x)_{\boldsymbol{U}}] \neq \underbrace{[1 - a_{\boldsymbol{U}} + (ac)_{\boldsymbol{L}},}_{a,\, ac \text{ not indep}}\ 1 - a_{\boldsymbol{L}} + (ac)_{\boldsymbol{U}}]$. We need to minimize $m = 1 - a + x$, but expressed using the lower bound of x from (3). So we need to minimize $m = 1 - a + \max\{a + b - 1, 0\}$. From the graph



we see $\dfrac{\partial m}{\partial a} \leq 0$ and $\dfrac{\partial m}{\partial b} \geq 0$, so we compute at the point $(a_{\boldsymbol{U}}, b_{\boldsymbol{L}})$, $m_{\boldsymbol{L}} = 1 - a_{\boldsymbol{U}} + \max\{a_{\boldsymbol{U}} + b_{\boldsymbol{L}} - 1,\ 0\} = \begin{cases} 1 - a_{\boldsymbol{U}} & \text{if } b_{\boldsymbol{L}} \leq 1 - a_{\boldsymbol{U}} \\ b_{\boldsymbol{L}} & \text{if } b_{\boldsymbol{L}} > 1 - a_{\boldsymbol{U}} \end{cases}$. And to maximize m we need to use the upper bound for x from (3). So we maximize $m = 1 - a + x = 1 - a + \min\{a, b\}$. From the graph



we see that $\dfrac{\partial m}{\partial a} \leq 0$ and $\dfrac{\partial m}{\partial b} \geq 0$, so we compute at the point $(a_{\boldsymbol{L}}, b_{\boldsymbol{U}})$, $m_{\boldsymbol{U}} = 1 - a_{\boldsymbol{L}} + \min\{a_{\boldsymbol{L}}, b_{\boldsymbol{U}}\} = \begin{cases} 1 - a_{\boldsymbol{L}} + b_{\boldsymbol{U}} & \text{if } b_{\boldsymbol{U}} \leq a_{\boldsymbol{L}} \\ 1 & \text{if } b_{\boldsymbol{U}} > a_{\boldsymbol{L}} \end{cases}$.

3.5 BOUNDS ON b

We next investigate different ways to obtain bounds on b. Let $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, and suppose we know m, but are not given any information about c. Hence we must compute $c \in [c_{\boldsymbol{L}},\ c_{\boldsymbol{U}}] = [(m + a_{\boldsymbol{L}} - 1)/a_{\boldsymbol{L}}, (m + a_{\boldsymbol{U}} - 1)/a_{\boldsymbol{U}}]$. Then $b \in [a_{\boldsymbol{L}} + m - 1,\ m]$, i.e., $b = [a_{\boldsymbol{L}}c_{\boldsymbol{L}},\ m]$, and $\Delta b = m - a_{\boldsymbol{L}}c_{\boldsymbol{L}} = m - (a_{\boldsymbol{L}} + m - 1) = 1 - a_{\boldsymbol{L}}$. Also note: $\Delta c = \dfrac{a_{\boldsymbol{U}} + m - 1}{a_{\boldsymbol{U}}} - \dfrac{a_{\boldsymbol{L}} + m - 1}{a_{\boldsymbol{L}}} = \dfrac{\Delta a}{a_{\boldsymbol{L}}a_{\boldsymbol{U}}}(1 - m)$. But since $a + m \geq 1$, for all consistent a, we have $\dfrac{\Delta a}{a_{\boldsymbol{L}}a_{\boldsymbol{U}}}(1 - m) \leq \dfrac{\Delta a}{a_{\boldsymbol{L}}a_{\boldsymbol{U}}}a_{\boldsymbol{L}} = \dfrac{\Delta a}{a_{\boldsymbol{U}}} = 1 - \dfrac{a_{\boldsymbol{L}}}{a_{\boldsymbol{U}}}$.

Now suppose we know c, but are not given any information about m. Hence we must compute $m \in [m_{\boldsymbol{L}},\ m_{\boldsymbol{U}}] = [1 - a_{\boldsymbol{U}}(1 - c),\ 1 - a_{\boldsymbol{L}}(1 - c)]$. Then $b \in [a_{\boldsymbol{L}}c,\ m_{\boldsymbol{U}}]$, and $\Delta b = m_{\boldsymbol{U}} - a_{\boldsymbol{L}}c = 1 - a_{\boldsymbol{L}}(1 - c) - a_{\boldsymbol{L}}c = 1 - a_{\boldsymbol{L}}$. Also, $\Delta m = 1 - a_{\boldsymbol{L}}(1 - c) - 1 + a_{\boldsymbol{U}}(1 - c) = \Delta a(1 - c)$. So we see that knowing m (and not c) gives a better $b_{\boldsymbol{U}}$, and knowing c (and not m) gives a better $b_{\boldsymbol{L}}$. But the Δb is the same in both cases.

Next suppose we know both m, c (exactly). Then using either material implication ($\gamma = 1$) or conditional ($\gamma = 0$) gives the same bounds $b \in [a_{\boldsymbol{L}}c,\ m]$. Now suppose we are given $m \in [m_{\boldsymbol{L}},\ m_{\boldsymbol{U}}]$ and $c \in [c_{\boldsymbol{L}},\ c_{\boldsymbol{U}}]$. Then inference with material implication gives $b_{\boldsymbol{L}} = a_{\boldsymbol{L}}c(a_{\boldsymbol{L}}, m_{\boldsymbol{L}}) = a_{\boldsymbol{L}} + m_{\boldsymbol{L}} - 1$, $b_{\boldsymbol{U}} = m_{\boldsymbol{U}}$ which gives $\Delta b = 1 - a_{\boldsymbol{L}} + \Delta m$. Bayesian inference gives $b_{\boldsymbol{L}} = a_{\boldsymbol{L}}c_{\boldsymbol{L}}$, $b_{\boldsymbol{U}} = m(a_{\boldsymbol{L}}, c_{\boldsymbol{U}}) = 1 - a_{\boldsymbol{L}}(1 - c_{\boldsymbol{U}})$, and $\Delta b = 1 - a_{\boldsymbol{L}}(1 - \Delta c) = 1 - a_{\boldsymbol{L}} + a_{\boldsymbol{L}}\Delta c$. And if we kept the best bounds from both, we would have $b_{\boldsymbol{L}} = \max\{a_{\boldsymbol{L}} + m_{\boldsymbol{L}} - 1,\ a_{\boldsymbol{L}}c_{\boldsymbol{L}}\}$, $b_{\boldsymbol{U}} = \min\{m_{\boldsymbol{U}},\ 1 - a_{\boldsymbol{L}}(1 - c_{\boldsymbol{U}})\}$.

A priori, any of these four bounds could be optimal. In practice, the values of $c_{\boldsymbol{L}}, m_{\boldsymbol{U}}$ would already be reset to the computed values if the original values were less optimal than the computed ones in the optimization/clipping phase. In this case we would just have $b \in [a_{\boldsymbol{L}}c_{\boldsymbol{L}}, m_{\boldsymbol{U}}]$. Then $\Delta b = m_{\boldsymbol{U}} - a_{\boldsymbol{L}}c_{\boldsymbol{L}}$. We can't simplify algebraically, but $m_{\boldsymbol{U}} \geq m(a_{\boldsymbol{L}}, c_{\boldsymbol{U}})$ and $c_{\boldsymbol{L}} \leq c(a_{\boldsymbol{L}}, m_{\boldsymbol{L}}) \leq c_{\boldsymbol{U}}$ give $\Delta b \geq m(a_{\boldsymbol{L}}, c_{\boldsymbol{U}}) - a_{\boldsymbol{L}}c_{\boldsymbol{L}} = (1 - a_{\boldsymbol{L}}) + a_{\boldsymbol{L}}\Delta c$, $\Delta b \geq m_{\boldsymbol{U}} - a_{\boldsymbol{L}}c(a_{\boldsymbol{L}}, m_{\boldsymbol{L}}) = (1 - a_{\boldsymbol{L}}) + \Delta m$.

Now we want to compute bounds on b when we have a gen-

iteration	$m_{\boldsymbol{L}}$ $1-a_1(1-c_0)$	$m_{\boldsymbol{U}}$ $1-a_0(1-c_1)$	$c_{\boldsymbol{L}}$ $1-(1-m_0)/a_0$	$c_{\boldsymbol{U}}$ $1-(1-m_1)/a_1$	$a_{\boldsymbol{L}}$ $(1-m_1)/(1-c_0)$	$a_{\boldsymbol{U}}$ $(1-m_0)/(1-c_1)$
start	0	1	0	1	0	1
1	$\boxed{0.5}$	1	$\underline{0}$	1	0	$\overline{1}$
2	$\underline{0.5}$	$\boxed{0.8}$	$\underline{0}$	0.8	0.2	$\overline{1}$
3	$\underline{0.5}$	$\overline{0.8}$	$1/6$	0.8	$\boxed{0.6}$	$\overline{1}$
4	$\underline{0.5}$	$\overline{0.8}$	$1/6$	0.75	$\underline{0.6}$	$\boxed{0.8}$
start	0	1	0	1	0	1
1	0	1	0	1	$\boxed{0.6}$	1
2	0.2	1	$\underline{0}$	1	0.6	$\boxed{0.8}$
3	$\boxed{0.5}$	1	$1/6$	1	$\underline{0.6}$	$\overline{0.8}$
4	$\underline{0.5}$	$\boxed{0.8}$	$1/6$	0.75	$\underline{0.6}$	$\overline{0.8}$
start	$\boxed{0.2}$	$\boxed{0.8}$	0	1	$\boxed{0.1}$	$\boxed{0.9}$
1	$\underline{0.2}$	$\overline{0.8}$	$\underline{0}$	$7/9$	0.2	$\overline{0.9}$
start	$\boxed{0.3}$	$\boxed{0.8}$	0	1	0	1
1	0.3	0.8	0	0.8	0.2	1

Figure 5: Example computations: Each bound that is manually set at an iteration (e.g., new information external to the system) is in a $\boxed{\text{box}}$. If a value is outside the current bounding box, then $\underline{x}$ means it was clipped up to the old bound $x_{\boldsymbol{L}}$ (bounded below), and $\overline{x}$ means it was clipped down to the old value $x_{\boldsymbol{U}}$ (bounded above). Note that if there are multiple ways to compute a bound, e.g., $m'_{\boldsymbol{U}}(a_{\boldsymbol{L}}, c_{\boldsymbol{U}})$, $m'_{\boldsymbol{U}}(a'_{\boldsymbol{L}}, c_{\boldsymbol{U}})$, $m'_{\boldsymbol{U}}(a_{\boldsymbol{L}}, c'_{\boldsymbol{U}})$, $m'_{\boldsymbol{U}}(a'_{\boldsymbol{L}}, c'_{\boldsymbol{U}})$, they will not always yield the same result. We must always compute with the most recent data, since the order that the updates come to us matters to the final answer.

eralized implication $A \xrightarrow{\gamma} B$. The equations in Lemma 2 are true for any B with $\boldsymbol{P}(B \mid A) = c$, but b is itself not functionally related to any of m, c, a. Consequently, we can only get bounds on b, even if we know all of m, c, a, x precisely. In this case, we use $\boldsymbol{P}(A \cap B) \leq \boldsymbol{P}(B) \leq \boldsymbol{P}(\overline{A} \cup B)$, to get the well-known (tight) bounds $x \leq b \leq m$, i.e., $b_{\boldsymbol{L}} = x$, $b_{\boldsymbol{U}} = m$. Using Lemma 4.ii, 4.iii, we can express these bounds in terms of g:

$$b_{\boldsymbol{L}} = (1-\gamma)ag + \gamma(a+g-1), \tag{5}$$
$$b_{\boldsymbol{U}} = (1-\gamma)ag + \gamma(a+g-1) + (1-a). \tag{6}$$

Of course, these formulas apply when we have precise probabilities for a, g. If we only have bounds for a and g, we compute $\frac{\partial}{\partial a} b_{\boldsymbol{L}}(a,g) > 0$, $\frac{\partial}{\partial g} b_{\boldsymbol{L}}(a,g) > 0$, and $\frac{\partial}{\partial a} b_{\boldsymbol{U}}(a,g) < 0$, $\frac{\partial}{\partial g} b_{\boldsymbol{U}}(a,g) > 0$. Using Lemma 9 we can compute $(b_{\boldsymbol{L}})_{\boldsymbol{L}} = x_{\boldsymbol{L}}$, $(b_{\boldsymbol{L}})_{\boldsymbol{U}} = x_{\boldsymbol{U}}$, $(b_{\boldsymbol{U}})_{\boldsymbol{L}} = m_{\boldsymbol{L}}$, $(b_{\boldsymbol{U}})_{\boldsymbol{U}} = m_{\boldsymbol{U}}$. 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:

$$b_{\boldsymbol{L}} = b_{\boldsymbol{L}}(a_{\boldsymbol{L}}, g_{\boldsymbol{L}}) = (1-\gamma)a_{\boldsymbol{L}}g_{\boldsymbol{L}} + \gamma(a_{\boldsymbol{L}} + g_{\boldsymbol{L}} - 1) \tag{7}$$
$$\begin{aligned} b_{\boldsymbol{U}} &= b_{\boldsymbol{U}}(a_{\boldsymbol{L}}, g_{\boldsymbol{U}}) \\ &= (1-\gamma)a_{\boldsymbol{L}}g_{\boldsymbol{U}} + \gamma(a_{\boldsymbol{L}} + g_{\boldsymbol{U}} - 1) + (1-a_{\boldsymbol{L}}). \end{aligned} \tag{8}$$

Whether $b_{\boldsymbol{L}}$ refers to the function in (5) or the value in (7) 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 $\Delta b = 1 - a_{\boldsymbol{L}}$, as noted above for the case of precise m, c.

4 IMPLICATION CHAINS WITH GENERALIZED IMPLICATION

Implications chains (i.e., *hypothetical syllogism*) are of particular interest for rule-based reasoning, hence we briefly summarize the properties of the generalized implication chains.

Suppose we have 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 we are given bounds, $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, on the initial antecedent A. We want to compute the bounds for the final consequent b^N. We first do this for arbitrary γ. Then we easily derive the bounds for material implication (*modus ponens*, *hypothetical syllogism*) and Bayesian inference by setting $\gamma = 0, 1$.

Theorem 11. *Given a sequence of generalized implications $B^i \xrightarrow{\gamma^i} B^{i+1}$, for $i = 0, 1, \ldots, N-1$, with $A = B^0$, and given the bounds $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$. Fix arbitrary γ, and suppose that $\gamma^i = \gamma$, with $g^i \in [g^i_{\boldsymbol{L}},\ g^i_{\boldsymbol{U}}]$, for $i < N$. Let $X^i = \gamma(1-g^i) + g^i$ and $Y^i = \gamma(1-g^i)$. Then the bounds at the*

end of the generalized implication chain are

$$b_{\boldsymbol{L}}^N = a_{\boldsymbol{L}} \prod_{i=0}^{N-1} X_{\boldsymbol{L}}^i - \sum_{i=0}^{N-1} \left(Y_{\boldsymbol{U}}^i \prod_{j=i+1}^{N-1} X_{\boldsymbol{L}}^j \right),$$

$$b_{\boldsymbol{U}}^N = 1 - Y_{\boldsymbol{L}}^{N-1} - (1 - X_{\boldsymbol{U}}^{N-1}) \left[a_{\boldsymbol{L}} \prod_{i=0}^{N-2} X_{\boldsymbol{U}}^i - \sum_{i=0}^{N-2} \left(Y_{\boldsymbol{L}}^i \prod_{j=i+1}^{N-2} X_{\boldsymbol{U}}^j \right) \right].$$

Proof. We first note that $X_{\boldsymbol{L}}^i = X^i(g_{\boldsymbol{L}}^i)$, $X_{\boldsymbol{U}}^i = X^i(g_{\boldsymbol{U}}^i)$, $Y_{\boldsymbol{L}}^i = Y^i(g_{\boldsymbol{U}}^i)$, $Y_{\boldsymbol{U}}^i = Y^i(g_{\boldsymbol{L}}^i)$. To prove the claim we give a straightforward induction on n. For the base case $n = 1$, simple substitution and algebra gives $b_{\boldsymbol{L}}^1 = a_{\boldsymbol{L}} X_{\boldsymbol{L}}^0 - Y_{\boldsymbol{U}}^0 = (1-\gamma) a_{\boldsymbol{L}} g_{\boldsymbol{L}}^0 + \gamma(a_{\boldsymbol{L}} + g_{\boldsymbol{L}}^0 - 1)$, which is just (7). Similarly, $b_{\boldsymbol{U}}^1 = 1 - a_{\boldsymbol{L}}(1 - X_{\boldsymbol{U}}^0) - Y_{\boldsymbol{L}}^0 = 1 - a_{\boldsymbol{L}} + (1-\gamma) a_{\boldsymbol{L}} g_{\boldsymbol{U}}^0 + \gamma(a_{\boldsymbol{L}} + g_{\boldsymbol{U}}^0 - 1)$, which is just (8).

For the inductive step, we get $b_{\boldsymbol{L}}^{n+1} = b_{\boldsymbol{L}}^n X_{\boldsymbol{L}}^n - Y_{\boldsymbol{U}}^n$
$= \left[a_{\boldsymbol{L}} \prod_{i=0}^{n-1} X_{\boldsymbol{L}}^i - \sum_{i=0}^{n-1} \left(Y_{\boldsymbol{U}}^i \prod_{j=i+1}^{n-1} X_{\boldsymbol{L}}^j \right) \right] X_{\boldsymbol{L}}^n - Y_{\boldsymbol{U}}^n$
$= a_{\boldsymbol{L}} \prod_{i=0}^{n} X_{\boldsymbol{L}}^i - \sum_{i=0}^{n} \left(Y_{\boldsymbol{U}}^i \prod_{j=i+1}^{n-1} X_{\boldsymbol{L}}^j \right) X_{\boldsymbol{L}}^n - Y_{\boldsymbol{U}}^n$
$= a_{\boldsymbol{L}} \prod_{i=0}^{n} X_{\boldsymbol{L}}^i - \sum_{i=0}^{n} \left(Y_{\boldsymbol{U}}^i \prod_{j=i+1}^{n} X_{\boldsymbol{L}}^j \right) X_{\boldsymbol{L}}^n$, as desired, and similarly for $b_{\boldsymbol{U}}^{N+1}$.

Recall that $b_{\boldsymbol{L}}^i = \cdots$, $b_{\boldsymbol{U}}^i = \cdots$ are automatically clipped to remain in $[0, 1]$, as per Definition 5, 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. □

Note that since $\gamma = 0 \Rightarrow X = c,\ Y = 0$, and $\gamma = 1 \Rightarrow X = 1,\ Y = 1 - m$, we can easily recover the classical cases below.

Corollary 12. If we are using Bayesian inference, then $\gamma = 0$. So $c^i \in [c_{\boldsymbol{L}}^i,\ c_{\boldsymbol{U}}^i]$, for $i < N$ are given, and the bounds at the end of the implication chain are

$$b_{\boldsymbol{L}}^N = a_{\boldsymbol{L}} \prod_{i=0}^{N-1} c_{\boldsymbol{L}}^i\,, \quad b_{\boldsymbol{U}}^N = 1 - \left(a_{\boldsymbol{L}} \prod_{i=0}^{N-1} c_{\boldsymbol{L}}^i \right) \left(1 - c_{\boldsymbol{U}}^{N-1}\right)$$

We note that for typical data (not all $c_{\boldsymbol{L}}^i = 1$) we see $b_{\boldsymbol{L}}^N \to 0$ and $b_{\boldsymbol{U}}^N \to 1$.

Corollary 13. If we are using material implications in our chain, we just set $\gamma = 1$ in the formulas above. So $m^i \in [m_{\boldsymbol{L}}^i,\ m_{\boldsymbol{U}}^i]$, for $i < N$, are given, and the bounds at the end of the implication chain are

$$b_{\boldsymbol{L}}^N = a_{\boldsymbol{L}} - \sum_{i=0}^{N-1} \left(1 - m_{\boldsymbol{L}}^i\right)\,, \qquad b_{\boldsymbol{U}}^N = m_{\boldsymbol{U}}^{N-1}.$$

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

Lemma 14. *Suppose we are given 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 we are given the bounds $a \in [a_{\boldsymbol{L}},\ a_{\boldsymbol{U}}]$, and the bounds $g^i \in [g_{\boldsymbol{L}}^i,\ g_{\boldsymbol{U}}^i]$, for $i < N$. Then the bounds for the end of the generalized implication chain can be computed using Theorem 11.*

Proof. We just use Lemma 4.vii to write all of the γ^i, g^i pairs in terms of a single fixed $\hat{\gamma}$ (of our choice), and the newly computed $\hat{g}^i$. Then apply Theorem 11. However, these bounds might not be optimal for the reasons discussed in Section 3.5. □

Lemma 15. *The generalized implication chain, $B^i \xrightarrow{\gamma} B^{i+1}$, for $i = 0, 1, \ldots, N-1$, from Theorem 11, can be replaced with a single implication $A \xrightarrow{\gamma} B$, where $A = B^0$, $B = B^N$, and the bounds $[b_{\boldsymbol{L}}^N,\ b_{\boldsymbol{U}}^N]$ are the same, whether computed from the implication chain or the single implication.*

Proof. Note that everything is specified except for the bounds on g. Since γ, $a_{\boldsymbol{L}}$, $b_{\boldsymbol{L}} = b_{\boldsymbol{L}}^N$, and $b_{\boldsymbol{U}} = b_{\boldsymbol{U}}^N$ are all fixed, equations (7), (8) give two linear equations each with one unknown. □

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 (8)).

(b) The lower bound $b_{\boldsymbol{L}}^N$ 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 $b_{\boldsymbol{U}}^N$ 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 $b_{\boldsymbol{U}}^N$.

(d) In the case $\gamma = 1$ (material implication), we compute $\Delta b^N = b_{\boldsymbol{U}}^N - b_{\boldsymbol{L}}^N = \Delta b^N = \Delta m^{N-1} + (1 - a_{\boldsymbol{L}}) + \sum_{i=0}^{N-2}(1 - m_{\boldsymbol{L}}^i)$. 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,

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

5 DISCUSSION

Our investigation of the properties of the generalized rule 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 implications 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. In general, one possible application of the γ parameter is to make if a function of the amount of data from which rules have been extracted: the more data is available, the closer the representation should go do the normatively correct Bayesian end point.

6 CONCLUSION AND FUTURE WORK

The goal of this paper was to investigate the relationship of probabilistic material implications and Bayesian conditionals, and for this purpose we utilized a generalization that subsumed both as special cases. We provided methods for obtaining various bounds from different rules and also proved a generalization for implication chains that determines the bounds for mixed rule-based inference chains. While the goal of this paper was not to speculate about interpretations of the γ parameter of the generalized rule, it would be interesting to see whether there is a natural interpretation of γ. We could only briefly allude to one way of thinking about it in terms of empirical support, i.e., the amount of data available for deriving probabilistic rules, but other interpretations are certainly possible and would be interesting to pursue.

References

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

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

Ivano Ciardelli and Adrian Ommundsen. Probabilities of conditionals: updating adams. *Noûs*, n/a (n/a). doi: https://doi.org/10.1111/nous.12437. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/nous.12437`.

Vincenzo Crupi and Andrea Iacona. Three ways of being non-material. *Studia Logica*, (1):1–47, 2021. doi: 10.1007/s11225-021-09949-y.

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/0888613X90900070`.

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.

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.

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`.

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`.

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

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`.

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.

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`.

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.

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.

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.