Lossy Inference Rules and their Bounds: a Brief Review

David Makinson and James Hawthorne

in A. Koslow, A. Buchsbaum (eds.), *The Road to Universal Logic*, Studies in Universal Logic, Springer, 2015, pp. 385-407.

Abstract This paper reviews results that have been obtained about bounds on the loss of probability occasioned by applying classically sound, but probabilistically unsound, Horn rules for inference relations. It uses only elementary finite probability theory without appealing to linear algebra, and also provides some new results, in the same spirit, on non-Horn rules. More specifically, it does the following: (1) recalls Adams' well-known sum bound for the rule Right∧+ and shows how it is inherited by the rules CM, CT and Left∨+; (2) draws attention to lesser known but tighter bounds for CM, CT and Left∨+ due respectively to Bourne & Parsons, Adams, and Gilio, and provides elementary verifications for those that were originally obtained using linear algebra; (3) shows that the sum bound for Right∧+ and the improved bounds for CM, CT and Left∨+ are all in a natural sense optimal, (4) distinguishes two kinds of loss for almost-Horn rules, 'distributed' and 'pointed'; and (5) finds bounds for both distributed and pointed loss, optimal in the distributed case, for the specific almost-Horn rules of disjunctive rationality (DR) and rational monotony (RM).

Keywords Uncertain inference · Probabilistic inference · Lossy rules · Preferential inference · Nonmonotonic reasoning · Horn rules · almost-Horn rules

Mathematics Subject Classification Primary 03B48 · Secondary 03B05 · 03B60 · 68T27

1 Introduction

It is a great pleasure to participate in the *Festschrift* for Jean-Yves Béziau. This paper is in the spirit of what he dubbed 'universal logic': it concerns the probabilistic assessment of rules for uncertain inference, and brings out close connections between quantitative and purely qualitative perspectives.

As is well known, one may model relations |~ of uncertain inference probabilistically, interpreting $a \mid\sim x$ as saying that $p(x|a) \geq t$ modulo a given probability function p and threshold value t taken as parameters. It is also well known that a single application of the inference rule of monotony (aka strengthening the antecedent) for an inference relation |~

Whenever $a \mid\sim x$, then $a \wedge b \mid\sim x$

can occasion a total loss of probability. No matter how close the conditional probability $p(x|a)$ (associated with the premise $a \mid\sim x$) may be to unity, so long as it doesn't equal unity the conditional probability $p(x|a \wedge b)$ (associated with the conclusion $a \wedge b \mid\sim x$) can be zero.

In contrast, some other rules do not lead to any loss of probability, an example being the rule of very cautious monotony (VCM):

Whenever $a \mid\sim b \wedge x$, then $a \wedge b \mid\sim x$.

The probability $p(x|a \wedge b)$ associated with its conclusion is always at least as high as the probability $p(b \wedge x|a)$ associated with its premise.

There are also some rules 'in between', whose application may lose some, but not all the probability of the premises; we may describe them as *lossy*. An important example is the rule of conjunction of consequents, also known as Right∧+:

Whenever $a \mid\sim x$ and $a \mid\sim y$, then $a \mid\sim x \wedge y$.

So long as the conditional probabilities associated with the premises $a \mid\sim x$ and $a \mid\sim y$ are 'sufficiently high', then that of the conclusion $a \mid\sim x \wedge y$ must be 'reasonably high', with an identifiable bound on how low it can fall.

This paper is about rules of the intermediate kind and their bounds. As well as Right∧+, we will consider cautious monotony (CM), cumulative transitivity (CT) and unrestricted Left∨+, whose definitions are recalled below. Some of their bounds have been established using linear algebra; we verify them all using only elementary finite probability theory, with some alternative proofs using material from non-monotonic logic. We also begin an investigation of bounds for non-Horn rules, focussing on those of disjunctive rationality (DR) and rational monotony (RM).

After section 2 on background concepts, all sections end with a short summary and all bounds are displayed in boxes. Readers may wish to consult them before deciding whether to plunge into details.

2 Background Concepts

We recall the notion of probabilistic soundness for Horn rules, the formulations of the specific Horn rules mentioned above, and the useful concept of sum soundness.

2.1 Probabilistic Soundness

Consider any n-premise Horn rule for inference relations, authorizing passage from a set of premises $a_i \mid\sim x_i$ ($i \leq n$) to conclusion $b \mid\sim y$, where the formulae are from classical propositional logic. Note that the premises and the conclusion of such a rule are not themselves formulae of classical logic, but are expressions $a \mid\sim x$ saying that the inference relation $\mid\sim$ holds between classical formulae a and x.

The rule is said to be *probabilistically sound* iff $p(y|b) \geq min\{p(x_i|a_i) : i \leq n\}$ for every probability function p. In other words iff, for every such function p and every $t \in [0,1]$, if $p(x_i|a_i) \geq t$ for all $i \leq n$ then also $p(y|b) \geq t$. Here, probability functions are understood as satisfying the usual (finitary) Kolmogorov postulates, and conditional probability $p(x|a)$ is defined in the standard way, putting $p(x|a) = p(a \wedge x)/p(a)$ when $p(a) \neq 0$. It will be convenient to let $p(x|a) = 1$ when $p(a) = 0$, rather than leave it undefined. In the limiting case that $n = 0$, the definition of probabilistic soundness is understood as requiring that $p(y|b) = 1$ for every probability function p.

This definition may be expressed more concisely if we write $a \mid\sim_{p,t} x$ for $p(x|a) \geq t$: a rule from premises $a_i \mid\sim x_i$ ($i \leq n$) to conclusion $b \mid\sim y$ is probabilistically sound iff $b \mid\sim_{p,t} y$ whenever $a_i \mid\sim_{p,t} x_i$ for all $i \leq n$. It may also be expressed in terms of improbability: the requirement

becomes $1-p(y|b) \leq max\{1-p(x_i|a_i)\colon i\leq n\}$, that is, $imp(y|b) \leq max\{imp(x_i|a_i)\}\colon i\leq n\}$ where we write $imp(\cdot|\cdot)$ for $1-p(\cdot|\cdot)$ as is convenient for expressions with a single probability function p.

We have already mentioned one example of a probabilistically sound Horn rule, namely very cautious monotony (VCM). Other important examples, recalled in detail in section 4.1 below, include reflexivity, left classical equivalence, right weakening, exclusive Left∨+, and weak Right∧+. Taken together with very cautious monotony they constitute the family O, which is nearly, but not quite, complete for probabilistically sound Horn rules.

It was shown in Hawthorne and Makinson [10] Observation 2.4 that probabilistically sound Horn rules are always preferentially sound, that is, they hold in the well-known system **P** of preferential inference. In turn, preferentially sound Horn rules are always classically sound (see e.g. Kraus, Lehmann & Magidor [11] or the textbook presentation Makinson [13]). But both converses fail. A notorious example of a classically sound Horn rule that is not preferentially sound is monotony (defined above). Salient examples of preferentially sound Horn rules that are not probabilistically sound are conjunction of conclusions (Right∧+: from $a \mid\sim x$, $a \mid\sim y$ to $a \mid\sim x\wedge y$), cautious monotony (CM: from $a \mid\sim x$, $a \mid\sim y$ to $a\wedge x \mid\sim y$), cumulative transitivity (CT: from $a \mid\sim x$, $a\wedge x \mid\sim y$ to $a \mid\sim y$) and unrestricted disjunction in the premises (Left∨+: from $a \mid\sim x$, $b \mid\sim x$ to $a\vee b \mid\sim x$).

For all of them, and others, the question thus arises: how much probability can be lost in their application? In other words, how *lossy* are they?

2.2 Sum Soundness

In answering this question, the notion of sum soundness has for long played an important role. A Horn rule $a_i \mid\sim x_i$ $(i\leq n)$ / $b \mid\sim y$ is said to be *sum sound* iff the improbability of y given b is always less than or equal to the sum of the improbabilities of the x_i given their respective a_i. Briefly, iff $imp(y|b) \leq \Sigma\{imp(x_i|a_i)\colon i\leq n\}$ for every probability function p. Sum soundness thus contrasts neatly with probabilistic soundness: *sum* simply replaces *max* in the improbability formulation of the latter given above. The notion of sum soundness may also be expressed rather less transparently in terms of probability rather than improbability, holding iff $p(y|b) \geq \Sigma\{p(x_i|a_i)\colon i\leq n\}-(n-1)$.

For $n = 2$, which is the case that will principally concern us, a Horn rule is sum sound iff $imp(y|b) \leq imp(x_1|a_1)+imp(x_2|a_2)$; and when $n = 1$, iff $imp(y|b) \leq imp(x_1|a_1)$. In the limiting case that $n = 0$, the definition is understood as requiring $imp(y|b) = 0$. In terms of probability, the requirements for sum soundness are: for $n = 2$, $p(y|b) \geq p(x_1|a_1)+p(x_2|a_2) -1$, for $n = 1$, $p(y|b) \geq p(x_1|a_1)$; for $n = 0$, $p(y|b) = 1$.

For zero and one-premise Horn rules, the two concepts of sum soundness and probabilistic soundness thus coincide, and are known also to coincide with preferential soundness as shown e.g. in Hawthorne and Makinson [10] Observation 4.3. When more than one premise is in play, the two notions diverge: probabilistic soundness immediately implies sum soundness since $max\{imp(x_i|a_i)\colon i\leq n\} \leq \Sigma\{imp(x_i|a_i)\colon i\leq n\}$; but not in general conversely.

Historically, the concept of sum soundness for Horn rules of inference seems first to have been articulated by Adams [1]. It is a very useful yardstick when comparing the loss of probability that may be incurred when applying inference rules – and of course, the less loss, the better. But the choice of the word ‘sound’ does not mean that the property is a certificate of acceptability: sum soundness does not play the same kind of normative role as does probabilistic soundness, in other words, zero loss. For a general discussion of the notion of a lossy rule of inference and its use in analysing the celebrated lottery and preface paradoxes, see Makinson [14].

3 Conjunction of Conclusions (Right∧+)

We show that the rule of conjunction of conclusions is sum sound and that, in a natural sense, this bound is optimal for it.

3.1 Sum Soundness of (Right∧+)

Like monotony, the rule of conjoining conclusions (from $a \mid\sim x$, $a \mid\sim y$ to $a \mid\sim x\wedge y$) is not probabilistically sound. But it is not difficult to show, as was already done by Adams [1], that it is nevertheless sum sound. That is, $imp(x\wedge y|a) \leq imp(x|a)+imp(y|a)$, equivalently $p(x\wedge y|a) \geq p(x|a)+p(y|a)-1$.

This follows immediately from a stronger full equality which one might call *exact sum soundness* for Right∧+: $imp(x\wedge y|a) = imp(x|a)+imp(y|a)-imp(x\vee y|a)$, and which may be established by a short chain of equalities using de Morgan and the well-known probabilistic version of the 'principle of inclusion and exclusion': $imp(x\wedge y|a) = p(\neg(x\wedge y)|a) = p(\neg x\vee\neg y|a) = p(\neg x|a)+p(\neg y|a)-p(\neg x\wedge\neg y|a) = imp(x|a)+imp(y|a)-imp(x\vee y)|a)$.

When expressed 'positively' in terms of probabilities, exact sum soundness says: $p(x\wedge y|a) = p(x|a)+p(y|a)-p(x\vee y|a)$, which is nicely (and curiously) symmetric with the improbability formulation. The positive version may also be established directly: by the 'principle of inclusion and exclusion' $p(x\vee y|a) = p(x|a)+p(y|a)-p(x\wedge y|a)$, and swapping terms gives $p(x\wedge y|a) = p(x|a)+p(y|a)-p(x\vee y|a)$ as desired. For easy reference, both formulations are displayed in the accompanying table 'Sum bound for Right∧+'.

Although exact sum soundness is more precise than plain sum soundness – indeed, being an equality, it is as precise as possible – we would not ordinarily call it a 'bound' on the rule Right∧+. For, to calculate $p(x|a)+p(y|a)-p(x\vee y|a)$ one needs to know, in general, not only the values of $p(x|a)$ and $p(y|a)$ corresponding to the two premises of the rule Right∧+, but also that of $p(x\vee y|a)$. If one is to use the language of bounds, one could perhaps say that exact sum soundness gives a soft bound for the probability of the conclusion of Right∧+, contrasting it with the (hard) bound provided by plain sum soundness.

In general, a *bound* (or hard bound for emphasis) for an n-premise Horn rule passing from premises $a_i \mid\sim x_i$ ($i\leq n$) to conclusion $b \mid\sim y$ is a function f: $[0,1]^n\rightarrow$ into [0,1], expressed in purely arithmetic terms, such that $p(y|b) \geq f(p(x_i|a_i)$: $i\leq n)$ so long as $0 \neq p(x_i|a_i) \neq 1$ for all $i \leq n$. The reason for the proviso is that it is convenient to allow exceptions for extreme values 0,1 of the $p(x_i|a_i)$, as we will do in sections 4 and 6 for certain bounds on the rules CM and Left∨+. A soft bound could admit more arguments to the function f. We consider only hard bounds, and call them simply bounds.

In the table on the sum bound for Right∧+, the top row gives the bound in terms of improbability and the second row in terms of probability. The second row also contains the general form of that bound, where r,s are understood to be the probabilities of the premises of Right∧+ and t is the probability of its conclusion. The same pattern will be followed in tables for other rules.

Sum bound for Right∧+
$imp(x\wedge y|a) \leq imp(x|a) + imp(y|a)$

$$p(x \wedge y|a) \geq p(x|a) + p(y|a) - 1$$
$$t \geq r+s-1$$

3.2 Optimality of the Sum Bound for Right∧+

Bounds may be looser or tighter, and we will say that a (hard) bound f for an n-premise Horn rule $a_i \mid\sim x_i$ ($i \leq n$) / $b \mid\sim y$ is *optimal* (for that rule) iff for all $r_1,\ldots,r_n \in (0,1)$ there is a probability function p and elements, a_i,x_i,b,y of its domain with each $p(x_i|a_i) = r_i$ and $p(y|b) = max\{0, f(r_1,\ldots,r_n)\}$. Again, as indicated, we allow exceptions for extreme values 0,1 of the $p(x_i|a_i)$.

In this sense, the bound $f(r,s) = r+s-1$ for the rule Right∧+ is optimal: for all $r,s \in (0,1)$ there is a probability function p and elements, a,x,y of its domain with $p(x|a) = r$, $p(y|a) = s$, and $p(x \wedge y|a) = \max\{0, r+s-1\}$. The idea of the verification is to choose p etc. in such a way as to force $p(x \wedge y|a)$ to be 0 or $r+s-1$ according as $r+s \leq 1$ or $r+s > 1$, and then apply exact sum soundness.

Specifically, take the domain of p to be 2^S (the set of all subsets of S) where $S = \{1,2,3\}$, and consider two cases. *Case* 1. Suppose $r+s \leq 1$. Put $p(\{1\}) = r$, $p(\{3\}) = s$, $p(\{2\}) = 1-(r+s)$. This sums to one and so extends to a probability function on 2^S. Put $a = \{1,2,3\}$, $x = \{1\}$, $y = \{3\}$. Then $p(x \wedge y|a) = p(a \wedge x \wedge y)/p(a) = 0 = max\{0, r+s-1\}$ as desired. *Case* 2. Suppose $r+s > 1$. Put $p(\{1\}) = 1-s$, $p(\{3\}) = 1-r$, $p(\{2\}) = 1-[(1-s)+(1-r)] = (r+s)-1$. This also sums to one and so extends to a probability function on 2^S. Put $a = \{1,2,3\}$, $x = \{1,2\}$, $y = \{2,3\}$. Then $p(x \wedge y|a) = p(a \wedge x \wedge y)/p(a) = p(\{2\}) = max\{0, r+s-1\}$ and the verification is complete.

In presenting this verification we have, for convenience, mixed the propositional and field-of-subsets ways of speaking of the domain of a probability function; but it should be clear how one could pedantically write out the proof in one or the other mode alone.

3.3 Summary on Lossiness for Right∧+

Thus, the loss of probability occasioned by an application of Right∧+ is known to be limited: the improbability in the conclusion is never more than the sum of the improbabilities in the two premises, and the bound is, in a natural sense, optimal. The former feature may be seen as providing a justification for monitored application of the rule in probabilistic contexts, preferably beginning from premises whose probability is well above the threshold that is taken as the norm for the context in which one is working.

In other words, the rule Right∧+ is 'lossy', rather like procedures for compressing the data for a digital image. The amount of loss is variable, but we have a relatively small maximal loss on each application, and can tolerate it if the initial images/premises are of high quality and we do not apply the compression/inference too often.

4 Inheriting the Sum Bound from Right ∧+

The bound for ∧+ prompts a search for bounds on other Horn rules that are preferentially but not probabilistically sound. In particular, cautious monotony CM, cumulative transitivity CT, and disjunction in the premises Left∨+, (all formulated in section 2.1) are all in that situation. How lossy are they? The first point to observe is that they too are sum sound. This was shown by Bourne & Parsons [5], [6], mainly using linear algebra; in this section we give elementary verifications.

4.1 Two Methods of Verification

In fact, there are two different kinds of elementary verification for the sum soundness of CM, CT, and Left∨+. One is by direct arithmetic calculation. The other is by taking a system of qualitative rules for probabilistic inference relations, and analysing a suitable derivation of the rule in the system augmented by Right∧+. We display both methods.

For some of the verifications by arithmetic calculation, we will be making use of the following 'product trumps addition' fact: for $r,s \in [0,1]$, $r{\cdot}s \geq r+s-1$. Indeed, more strongly, if $r,s < 1$ then $r{\cdot}s > r+s-1$ while in the limiting case that $r = 1$ or $s = 1$, $r{\cdot}s = r+s-1$. The limiting case is immediate. To check the principal case, suppose $r{\cdot}s \leq r+s-1$. Then $1-s \leq r-r{\cdot}s = r{\cdot}(1-s)$ which implies $r = 1$ or $s = 1$.

For the verifications by analysis of derivations, we refer to Hawthorne's system **O**, all of whose rules are probabilistically sound. We recall from [8], [10] that it is made up of a zero-premise rule of reflexivity, one-premise rules of left logical equivalence and right weakening and a pair of rather special two-place rules, exclusive Left∨+ and weak Right∧+, as follows:

$a \mid\!\sim a$	(reflexivity)
Whenever $a \mid\!\sim x$ and $a \dashv\models b$, then $b \mid\!\sim x$	(LCE: left classical equivalence)
Whenever $a \mid\!\sim x$ and $x \models y$, then $a \mid\!\sim y$	(RW: right weakening)
Whenever $a \mid\!\sim x{\wedge}y$, then $a{\wedge}x \mid\!\sim y$	(VCM: very cautious monotony)
Whenever $a \mid\!\sim x$, $b \mid\!\sim x$ and $a \models \neg b$, then $a{\vee}b \mid\!\sim x$	(XOR: exclusive Left∨+)
Whenever $a \mid\!\sim x$ and $a{\wedge}\neg y \mid\!\sim y$, then $a \mid\!\sim x{\wedge}y$	(WAND: weak Right∧+).

Here $\models$ (in RW and XOR) is classical logical consequence and $\dashv\models$ (in LCE) is classical equivalence. They are auxiliary relations: notwithstanding their presence, LCE and RW are regarded as one-premise rules, since only one premise contains $\mid\!\sim$. Weak Right∧+ (WAND) is strictly weaker than plain Right∧+ (whenever $a \mid\!\sim x$ and $a \mid\!\sim y$, then $a \mid\!\sim x{\wedge}y$). Exclusive Left∨+ (XOR) differs from unrestricted Left∨+ (whenever $a \mid\!\sim x$ and $b \mid\!\sim x$, then $a{\vee}b \mid\!\sim x$) in that it is subject to the condition that a is classically inconsistent with b, and in the context of system **O** it is strictly weaker than the unrestricted version. Given the other rules of **O**, XOR may equivalently be expressed as: whenever $a{\wedge}b \mid\!\sim x$, $a{\wedge}\neg b \mid\!\sim x$, then $a{\vee}b \mid\!\sim x$. In that form it is also known as weak Left∨+ or weak OR (WOR); for more see [8], [10].

Replacing the weakened rules exclusive Left∨+ and weak Right∧+ by their plain counterparts gives us the system **P** of preferential consequence. Indeed, just replacing weak Right∧+ by plain Right∧+ suffices for that. Thus we know that CM, CT, Left∨+, being preferentially valid, are derivable from **O** supplemented by Right∧+. That alone does not imply that any lower bound for Right∧+ also serves as such for those rules, for the derivation could apply Right∧+ several times. But, as we will see, inspection of standard derivations in the literature reveals that in each of them ∧+ is applied only once, so that the bound is indeed inherited.

4.2 Cautious Monotony (CM)

The sum bound for CM may be obtained from that for Right∧+ by elementary arithmetic as follows. We want to show that $p(y|a{\wedge}x) \geq p(x|a)+p(y|a)-1$. By the bound for Right∧+, we know that $p(x{\wedge}y|a) \geq p(x|a)+p(y|a)-1$, so we need only show that $p(y|a{\wedge}x) \geq p(x{\wedge}y|a)$. In the limiting case that $p(a{\wedge}x) = 0$, $p(y|a{\wedge}x) = 1$ and we are done. In the case that $p(a{\wedge}x) \neq 0$ we also have

$p(a) \neq 0$ so that $p(y|a\wedge x) = p(a\wedge x\wedge y)/p(a\wedge x)$ while $p(x\wedge y|a) = p(a\wedge x\wedge y)/p(a)$, and we conclude by noting that $p(a) \geq p(a\wedge x)$.

Sum bound for CM
$imp(y\|a\wedge x) \leq imp(x\|a) + imp(y\|a)$
$p(y\|a\wedge x) \geq p(x\|a) + p(y\|a) - 1$ $t \geq r+s-1$

For a verification of the same bound for CM using logical means, we need a suitable notation. For any $\varepsilon \in [0,1]$, write $a \mid\!\sim_{\varepsilon} x$ as shorthand for $imp(x|a) \leq \varepsilon$, that is, $p(x|a) \geq 1-\varepsilon$, and note that since the rules of **O** are probabilistically sound we may apply them taking the index of the conclusion to be the maximum of the indices of the premises. In particular, this means that for the zero-premise rule of reflexivity the index is 0 and for the one-premise rules the index stays unchanged. For applications of Right∧+ the index goes up from $\varepsilon_1,\varepsilon_2$ to $\varepsilon_1+\varepsilon_2$, as already established.

With these indices we decorate a very short derivation of CM in **O**-plus-Right∧+. Given $a \mid\!\sim_{\varepsilon 1} x$ and $a \mid\!\sim_{\varepsilon 2} y$ we need to get $a\wedge x \mid\!\sim_{\varepsilon 1+\varepsilon 2} y$. From the two assumptions we have by Right∧+ that $a \mid\!\sim_{\varepsilon 1+\varepsilon 2} x\wedge y$ and so $a\wedge x \mid\!\sim_{\varepsilon 1+\varepsilon 2} y$ by VCM and we are done.

4.3 Cumulative Transitivity (CT)

Recall that CT authorizes passage from $a \mid\!\sim x$ and $a\wedge x \mid\!\sim y$ to $a \mid\!\sim y$; it is thus a partial converse of CM. We can give an arithmetic verification of its sum soundness without even using that for Right∧+. For, as already observed by Adams [3] (section 6.6, page 128) we have $p(y|a) \geq p(x\wedge y|a) = p(x|a)\cdot p(y|a\wedge x)$, and by substitution in the 'product trumps sum' principle mentioned above, $p(x|a)\cdot p(y|a\wedge x) \geq p(x|a)+p(y|a\wedge x)-1$. Putting these together, $p(y|a) \geq p(x|a)+p(y|a\wedge x)-1$ as desired. Later we will see how this arithmetic argument may be refined to give an improved bound for CT.

Sum bound for CT
$imp(y\|a) \leq imp(x\|a) + imp(y\|a\wedge x)$
$p(y\|a) \geq p(x\|a) + p(y\|a\wedge x) - 1$ $t \geq r+s-1$

For a 'logical' verification of the bound, we analyse a standard derivation of CT in the system **O**-plus-Right∧+. Suppose $a \mid\!\sim_{\varepsilon 1} x$ and $a\wedge x \mid\!\sim_{\varepsilon 2} y$; we need to get $a \mid\!\sim_{\varepsilon 1+\varepsilon 2} y$. By reflexivity and right weakening, $a\wedge\neg x \mid\!\sim_{0} \neg x\vee y$. On the other hand, $a\wedge x \mid\!\sim_{\varepsilon 2} y$ gives $a\wedge x \mid\!\sim_{\varepsilon 2} \neg x\vee y$ by right weakening. Combining these by XOR and LCE we have $a \mid\!\sim_{\varepsilon 2} \neg x\vee y$. Applying Right∧+ to that and $a \mid\!\sim_{\varepsilon 1} x$ gives $a \mid\!\sim_{\varepsilon 1+\varepsilon 2} x\wedge(\neg x\vee y)$, and so by right weakening $a \mid\!\sim_{\varepsilon 1+\varepsilon 2} y$ as desired.

Interestingly, while the arithmetic verification did not appeal to the sum-soundness of Right∧+, this logical one does apply that rule, and the application cannot be dispensed with since CT is not probabilistically sound and thus not derivable in the system **O** alone.

4.4 Disjunction in the Premises (Left∨+)

The sum soundness of Left∨+ was obtained by Bourne & Parsons [6], using linear programming. It can also be given a direct elementary verification, as we now show. As in the case of CT, we do not need to use the sum soundness of Right∧+.

We first note that $p(x|a) \leq p(x\vee\neg a|a\vee b)$ and $p(x|b) \leq p(x\vee a|a\vee b)$. For the former, it suffices to show $1-p(x\vee\neg a|a\vee b) \leq 1-p(x|a)$ as follows:

$$1-p(x\vee\neg a|a\vee b) = p(\neg x\wedge a|a\vee b) = p(\neg x|a)\cdot p(a|a\vee b) \leq 1-p(x|a).$$

For the latter, it likewise suffices to show $1-p(x\vee a|a\vee b) \leq 1-p(x|b)$ by the following chain:

$$1-p(x\vee a|a\vee b) = p(\neg x\wedge\neg a|a\vee b) = p(\neg x\wedge\neg a\wedge b|a\vee b) \leq p(\neg x\wedge b|a\vee b)$$
$$= p(\neg x|b)\cdot p(b|a\vee b) \leq 1-p(x|b).$$

Thus $p(x|a)+p(x|b) \leq p(x\vee\neg a|a\vee b)+p(x\vee a|a\vee b)$. But the right side of this inequality equals $1+p(x|a\vee b)$, as shown by the chain:

$$1 = p((x\vee\neg a)\vee(x\vee a)|a\vee b) = p(x\vee\neg a|a\vee b)+p(x\vee a|a\vee b)-p((x\vee\neg a)\wedge(x\vee a)|a\vee b)$$
$$= p(x\vee\neg a|a\vee b)+p(x\vee a|a\vee b)-p(x|a\vee b).$$

Putting these together, $p(x|a)+p(x|b) \leq 1+p(x|a\vee b)$, that is $p(x|a\vee b) \geq p(x|a)+p(x|b)-1$, as desired.

Sum bound for Left∨+
$imp(x\|a\vee b) \leq imp(x\|a) + imp(x\|b)$
$p(x\|a\vee b) \geq p(x\|a) + p(x\|b) - 1$ $t \geq r+s-1$

For a ‘logical’ verification of the same bound, we analyse a standard derivation of Left∨+ in the system **O**-plus-Right∧+ (with Right∧+ applied once). Suppose $a \mid\!\sim_{\varepsilon 1} x$ and $b \mid\!\sim_{\varepsilon 2} x$. We need to show that $a\vee b \mid\!\sim_{\varepsilon 1+\varepsilon 2} x$. Applying LCE to the first supposition, $(a\vee b)\wedge a \mid\!\sim_{\varepsilon 1} x$, so $(a\vee b)\wedge a \mid\!\sim_{\varepsilon 1} x\vee\neg a$ (RW); also $(a\vee b)\wedge\neg a \mid\!\sim_{0} x\vee\neg a$ (reflexivity and RW); thus $a\vee b \mid\!\sim_{\varepsilon 1} x\vee\neg a$ (XOR and LCE). Similarly, applying LCE to the second supposition, $(a\vee b)\wedge b \mid\!\sim_{\varepsilon 2} x$, so $(a\vee b)\wedge b \mid\!\sim_{\varepsilon 2} x\vee a$ (right weakening); also $(a\vee b)\wedge\neg b \mid\!\sim_{0} x\vee a$ (reflexivity and RW); thus $a\vee b \mid\!\sim_{\varepsilon 2} x\vee a$ (XOR and LCE). Putting these together by Right∧+ we have $a\vee b \mid\!\sim_{\varepsilon 1+\varepsilon 2} (x\vee a)\wedge(x\vee\neg a)$ so by RW $a\vee b \mid\!\sim_{\varepsilon 1+\varepsilon 2} x$ as desired.

4.5 Summary on Inheritance of the Sum Bound

The sum bound for Right∧+ is thus inherited by the rules CM (cautious monotony), CT (cumulative transitivity) and Left∨+ (disjunction of premises), as may be shown by two different kinds of argument. One proceeds by analysing the derivations of those rules in the system **O**-plus-Right∧+ for qualitative inference relations. Each derivation necessarily makes use of Right∧+, but with only one application required. The sum soundness of Right∧+ is then invoked, which is why we speak of inheritance of the property. The other kind of argument is directly arithmetic. Under this method of verification, appeal to the sum soundness of Right∧+ was made only in the case of CM so, from this more refined perspective the fortunes

of CT and Left$\vee$+ are actually self-earned rather than inherited. In the next section, we will see how the arithmetic arguments may be refined to yield improved bounds for CM, CT, Left$\vee$+.

5 Improved Bounds for CM, CT, Left$\vee$+

For all three derived rules, the sum bound may be tightened, and in this section we show how it may be done. We have reached the limit of the 'logical' method, and our arguments for the improved bounds are all arithmetic, although still elementary. They have a common feature: they all make use of the product or division of conditional probabilities, whereas the verifications of sum soundness used only addition and subtraction.

5.1 Improved Bound for CM

Using linear algebra, Bourne & Parsons [5], [6] established an improved bound for CM: when $p(x|a) \neq 0$, then $imp(y|a\wedge x) \leq [imp(x|a)+imp(y|a)] / p(x|a)$. In other words, it divides the sum bound, already known to hold for CM, by $p(x|a)$. Equivalently, in terms of probabilities: when $p(x|a) \neq 0$, then $p(y|a\wedge x) \geq [p(x|a)+p(y|a)-1] / p(x|a)$. We call this the 'divided sum' bound for CM, and write its general form as $t \geq (r+s-1)/r$, with the understanding that r is the probability $p(x|a)$ of the 'major' premise $a \mid\sim x$.

Clearly, whenever $p(x|a) < 1$, this is better than the plain sum bound, and in the limiting case that $p(x|a) = 1$ they are equal. When $p(x|a) = 0$ the quotient is undefined, but then $p(a\wedge x) = 0$ and thus, using the limiting case convention in the definition of conditional probability, $p(y|a\wedge x) = 1$ and $imp(y|a\wedge x) = 0$.

We verify the improved bound by elementary means. Recall that CM authorizes passage from $a \mid\sim x$ and $a \mid\sim y$ to $a\wedge x \mid\sim y$. Suppose $p(x|a) \neq 0$. We may assume wlog that $p(a\wedge x) \neq 0$, for otherwise $p(y|a\wedge x) = 1$ and we are done. Hence $p(y|a\wedge x) = p(a\wedge x\wedge y)/p(a\wedge x) = p(x\wedge y|a)/p(x|a)$, and we can replace the top $p(x\wedge y|a)$ by the sum bound for Right$\wedge$+ to get $p(y|a\wedge x) \geq [p(x|a)+p(y|a)-1] / p(x|a)$ as desired.

Improved bound for CM When $p(x\|a) \neq 0$:
$imp(y\|a\wedge x) \leq [imp(x\|a)+imp(y\|a)] / p(x\|a)$
$p(y\|a\wedge x) \geq [p(x\|a)+p(y\|a)-1] / p(x\|a)$ $t \geq (r+s-1)/r$

The divided sum bound is optimal for CM, in the sense defined in section 2. That is, for any $r,s \in (0,1)$ there is a probability function p and elements, a,x,y of its domain with $p(x|a) = r$, $p(y|a) = s$, and $p(y|a\wedge x) = max(0, (r+s-1)/r)$. The verification is quite similar to that of the optimality of sum soundness for Right$\wedge$+ in section 3.2. Let $r,s \in (0,1)$, put $S = \{1,2,3\}$ as before, and again consider two cases. *Case* 1. Suppose $r+s \leq 1$. Put p,a,x,y as before. Then $p(y|a\wedge x) = p(a\wedge x\wedge y)/p(a\wedge x) = 0 \geq (r+s-1)/r$ as desired. *Case* 2. Suppose $r+s > 1$. Put p,a,x,y as before. Then $p(y|a\wedge x) = p(a\wedge x\wedge y)/p(a\wedge x) = p(\{2\})/p(\{1,2\}) = (r+s-1)/r$ and the verification is complete.

5.2 Improved Bound for CT

Recall again that CT authorizes passage from $a \mid\sim x$ and $a\wedge x \mid\sim y$ to $a \mid\sim y$. In section 4.3 we noted that $p(y|a) \geq p(x\wedge y|a) = p(x|a)\cdot p(y|a\wedge x)$, which is already a bound for CT. We also noted in the same section that this is at least as good as the sum bound, since by the 'product trumps sum' principle we have $p(x|a)\cdot p(y|a\wedge x) \geq p(x|a)+p(y|a\wedge x)-1$. But that principle tells us a little more; it says that the last inequality is strict except when $p(x|a) = 1$ or $p(y|a\wedge x) = 1$. So $p(y|a) > p(x|a)+p(y|a\wedge x)-1$ except when one of $p(x|a)$, $p(y|a\wedge x)$ equals 1.

By how much does the left exceed the right? We can calculate the improvement as follows. Let $p(x|a) = (1-\varepsilon_1)$ and $p(y|a\wedge x) = (1-\varepsilon_2)$. Then $p(x|a)\cdot p(y|a\wedge x) = (1-\varepsilon_1)\cdot(1-\varepsilon_2) = [1-(\varepsilon_1+\varepsilon_2)]+(\varepsilon_1\cdot\varepsilon_2) = [(1-\varepsilon_1)+(1-\varepsilon_2)-1]+(\varepsilon_1\cdot\varepsilon_2) = [p(x|a)+p(y|a\wedge x)-1]+[(1-p(x|a))\cdot(1-p(y|a\wedge x))]$. The part in the left square parentheses is the sum bound, and the part in the right square parentheses is the improvement. Its general form is $t \geq r\cdot s = (r+s-1)+[(1-r)\cdot(1-s)]$.

Interestingly, although the rule CT is not symmetric around its two premises, the bound is nevertheless symmetric around their probabilities – as indeed was the sum bound. The improved bound can be also expressed more concisely in terms of improbabilities: $imp(y|a) \leq imp(x|a)+imp(y|a\wedge x)-imp(x|a)\cdot imp(y|a\wedge x)$.

Improved bound for CT
$imp(y\|a) \leq imp(x\|a)+imp(y\|a\wedge x)-imp(x\|a)\cdot imp(y\|a\wedge x)$
$p(y\|a) \geq p(x\|a)\cdot p(y\|a\wedge x) = [p(x\|a)+p(y\|a\wedge x)-1]+[(1-p(x\|a))\cdot(1-p(y\|a\wedge x)]$ $t \geq r\cdot s = (r+s-1)+[(1-r)\cdot(1-s)]$

The bound is optimal for CT. That is, for any $r,s \in (0,1)$ there is a probability function p and elements a,x,y of its domain with $p(x|a) = r$, $p(y|a\wedge x) = s$, such that $p(y|a) = r\cdot s$. To verify this, let $r,s \in (0,1)$, again take $S = \{1,2,3\}$, put $p(\{1\}) = r\cdot s$, $p(\{2\}) = r-r\cdot s$, $p(\{3\}) = 1-r$. These sum to 1 and so determine a probability function on 2^S. Put $a = S$, $x = \{1,2\}$, $y = \{1\}$. Then $p(x|a) = p(a\wedge x)/p(a) = r$, $p(y|a\wedge x) = p(a\wedge x\wedge y)/p(a\wedge x) = r\cdot s/r = s$, while $p(y|a) = p(a\wedge y)/p(a) = r\cdot s$ and we are done.

5.3 Comparison of the Improved Bounds for CM and CT

What is the relationship between the loss levels of the rules Right∧+, CM, and CT? In general terms, one can say that Right∧+ is lossier than either of CM and CT, which are incomparable in this respect.

To verify this, first recall that sum soundness is an *optimal* bound for Right∧+ and also a bound for CM and CT; so that Right∧+ is at least as lossy as the latter two. We also know that CM satisfies the 'divided sum' bound $(r+s-1)/r$, and that CT satisfies the 'product of premises' bound $r\cdot s$. So it will suffice to show that neither Right∧+ nor CM satisfies a product of premises bound, and that neither Right∧+ nor CT satisfies a divided sum bound.

For the former, we need to check that it can happen that for each of the rules Right∧+ and CM, the conclusion of the rule has probability *less* than $r\cdot s$ where r,s are the probabilities of the two premises of the rule. For a simple example covering both rules, let p be any probability function with each $p(\pm x\wedge\pm y) > 0$ and put $a = \neg(x\wedge y)$. With r,s being the values $p(x|a)$, $p(y|a)$ of

the premises of the rule, we have $r, s > 0$ so $r \cdot s > 0$, while the values $p(x \wedge y|a)$, $p(y|a \wedge x)$ of the respective conclusions of the two rules both equal 0.

For the latter, we need to check that it can happen that for each of the rules Right∧+ and CT, the conclusion of the rule has probability less than $(r+s-1)/r$ where $r = p(x|a)$ is the probability of the premise $a \mid\sim x$ (common to both rules) and s is the probability of the other premise of the rule, that is, $s = p(y|a)$ in the case of Right∧+ and $s = p(y|a \wedge x)$ in the case of CT. An example covering Right∧+ may be obtained as follows: let $S = \{1,2,3,4\}$ and $p(\{i\}) = 1/4$ for each $i \in S$. Put $a = S$, $x = \{1,2,3\}$, $y = \{3,4\}$. Then $r = p(x|a) = 3/4$, $s = p(y|a) = 1/2$, $p(x \wedge y|a) = 1/4$, so that $p(x \wedge y|a) < 1/3 = (r+s-1)/r$. For CT we take the same values for S,p,a,x but put $y = \{2,3\}$. Then $r = p(x|a) = 3/4$, $s = p(y|a \wedge x) = 2/3$, $p(y|a) = 1/2$, so that $p(y|a) < 5/9 = (r+s-1)/r$.

This relationship between the bounds for Right∧+, CM and CT is what one might have anticipated given the qualitative connections between those rules. While each of CM and CT is derivable in the system **O**-plus-Right∧+, neither Right∧+ nor CM is derivable in **O**-plus-CT, nor are Right∧+ or CT derivable in **O**-plus-CM, as shown in Hawthorne and Makinson [10] Observation 4.1. To be sure, this consonance is a matter of confirmed expectation rather than implication since, apart from anything else, Hawthorn's system **O** for probabilistic inference between Horn rules is known not to be complete, as was shown by Paris and Simmonds [15] with a brief overview in Hawthorne [9].

5.4 Improved Bound for Left∨+

Adams [2] (table I and appendix) gave an intricate and indirect proof that Left∨+ also satisfies the product bound $p(x|a \vee b) \geq p(x|a) \cdot p(x|b)$ and thus, in terms of improbability, $imp(x|a \vee b) \leq imp(x|a) + imp(x|b) - imp(x|a) \cdot imp(x|b)$. This is already an improvement on the plain sum bound on Left∨+ derived in Section 4.4 from the sum bound on Right∧+. The same result can also be obtained from the product bound for CT by a careful analysis of the derivation of Left∨+ from **O**-plus-CT that is given in Hawthorne and Makinson [10] (appendix, fact 4.1.4).

But a better bound was obtained by Gilio [7] (page 23) using linear algebra applied to de Finetti's theory of coherent probability assessments. His bound is:

$$p(x|a \vee b) \geq p(x|a) \cdot p(x|b) \,/\, [p(x|a) + p(x|b) - p(x|a) \cdot p(x|b)]$$

whenever this ratio is well-defined, that is, when either $p(x|a) \neq 0$ or $p(x|b) \neq 0$. This is the product bound divided by $p(x|a) + p(x|b) - p(x|a) \cdot p(x|b)$, and has the symmetric general form $r \cdot s \,/\, (r+s-r \cdot s)$. By the 'product trumps sum' principle, the divisor is less than 1 whenever both $p(x|a)$ and $p(x|b)$ are less than 1, so this bound for Right∨+ is better than the simple product $p(x|a) \cdot p(x|b)$ except in the limiting case that one of $p(x|a)$, $p(x|b)$ equals 1, in which case the two bounds are equal. As Gilio notes (page 26), his bound may likewise be written using improbabilities, in the rather more complex expression also recorded in the display.

Gilio's Improved bound for Left∨+ When either $p(x\|a) \neq 0$ or $p(x\|b) \neq 0$
$p(x\|a \vee b) \geq p(x\|a) \cdot p(x\|b) \,/\, [p(x\|a) + p(x\|b) - p(x\|a) \cdot p(x\|b)]$ $r \cdot s \,/\, (r+s-r \cdot s)$
$imp(x\|a \vee b) \leq [imp(x\|a) + imp(x\|b) - 2(imp(x\|a) \cdot imp(x\|b))] \,/\, 1-(imp(x\|a) \cdot imp(x\|b))$

We show how Gilio's bound for Left$\vee$+ may be verified without either linear algebra or detour through de Finetti's rather non-standard theory of coherent probability assessments. The argument is elementary but rather intricate; readers more interested in the result than its proof may wish to skip to the next section.

First, we may assume wlog that $p(a\vee b) \neq 0$, for otherwise $p(x|a\vee b) = 1$ and the Gilio inequality holds trivially. Indeed, we may assume wlog that $p(a\vee b) = 1$. For, given $p(a\vee b) \neq 0$, if LHS < RHS for a probability function p then, putting $q = p_{a\vee b}$, we have $q(x|c) = p(x|c)$ for all $c \vDash a\vee b$ so that for q we have LHS < RHS while $q(a\vee b) = 1$.

Thus, under the assumption $p(a\vee b) = 1$, we need only show $p(x) \geq p(x|a)\cdot p(x|b) / [p(x|a)+p(x|b)-p(x|a)\cdot p(x|b)]$, that is:

(1) $$p(x)\cdot[p(x|a)+p(x|b)-p(x|a)\cdot p(x|b)] \geq p(x|a)\cdot p(x|b).$$

Again, we dispose of some limiting cases. When $p(a) = 0$ then (1) holds. For suppose $p(a) = 0$. Then $p(x|a) = 1$ so RHS(1) = $p(x|b)$; also $p(b) = p(a\vee b)+p(a\wedge b)-p(a) = p(a\vee b) = 1$, so that $p(x) = p(x\wedge b)+p(x\wedge\neg b) = p(x\wedge b) = p(x|b)$, and thus also LHS(1) = $p(x|b)$. Similarly, (1) holds when $p(b) = 0$. So we may assume wlog that $p(a) \neq 0 \neq p(b)$. Under that assumption, when $p(x) = 0$ we have LHS(1) = 0 = RHS(1), so we may further suppose that $p(x) \neq 0$.

Next, we restate the problem. Multiplying both sides of (1) by the non-zero $p(a)\cdot p(b)$, it suffices to show:

$$p(x)\cdot p(a)\cdot p(b)[p(x|a)+p(x|b)-p(x|a)\cdot p(x|b)] \geq p(x|a)\cdot p(x|b)\cdot p(a)\cdot p(b).$$

Distributing, turning the left minus into a right plus, and eliminating conditional probabilities, it thus suffices to show:

(2) $$p(x)\cdot[p(a\wedge x)\cdot p(b) + p(b\wedge x)\cdot p(a)] \geq p(a\wedge x)\cdot p(b\wedge x)\cdot p(x) + p(a\wedge x)\cdot p(b\wedge x).$$

To complete the verification, we show that in (2), the left and right sides may be expressed as $\alpha_1+\alpha_2+\alpha_3$ and $\beta_1+\beta_2+\beta_3$ respectively, with each $\alpha_i \geq \beta_i$. We make use of the fact that since $p(a\vee b) = 1$, we can 'get rid' of $p(x)$ by the equality $p(x) = p((a\vee b)\wedge x) = p((a\wedge x)\vee(b\wedge x)) = p(a\wedge x)+p(b\wedge x)-p(a\wedge b\wedge x)$.

Substituting for $p(x)$ on the left, distributing and regrouping, LHS(2) = $\alpha_1+\alpha_2+\alpha_3$ where:

$$\alpha_1 = p(a\wedge x)\cdot p(b)\cdot[p(a\wedge x)-p(a\wedge b\wedge x)]$$
$$\alpha_2 = p(b\wedge x)\cdot p(a)\cdot[p(b\wedge x)-p(a\wedge b\wedge x)]$$
$$\alpha_3 = p(a\wedge x)\cdot p(b)\cdot p(b\wedge x) + p(b\wedge x)\cdot p(a)\cdot p(a\wedge x).$$

Substituting for $p(x)$ in the first term $p(a\wedge x)\cdot p(b\wedge x)\cdot p(x)$ on the right of (2) and distributing transforms that term into:

$$p(a\wedge x)\cdot p(b\wedge x)\cdot p(a\wedge x) + p(a\wedge x)\cdot p(b\wedge x)\cdot p(b\wedge x) - p(a\wedge x)\cdot p(b\wedge x)\cdot p(a\wedge b\wedge x)$$

so that, by subtracting and adding an additional $p(a\wedge x)\cdot p(b\wedge x)\cdot p(a\wedge b\wedge x)$, distributing, and finally adding the second term from the right of (2), we have RHS(2) = $\beta_1+\beta_2+\beta_3$ where:

$$\beta_1 = p(a\wedge x)\cdot p(b\wedge x)\cdot[p(a\wedge x) - p(a\wedge b\wedge x)]$$
$$\beta_2 = p(a\wedge x)\cdot p(b\wedge x)\cdot[p(b\wedge x) - p(a\wedge b\wedge x)]$$
$$\beta_3 = p(a\wedge x)\cdot p(b\wedge x)\cdot p(a\wedge b\wedge x) + p(a\wedge x)\cdot p(b\wedge x).$$

Clearly, since $p(b) \geq p(b\wedge x)$ and $p(a) \geq p(a\wedge x)$ we have $\alpha_1 \geq \beta_1$ and $\alpha_2 \geq \beta_2$. It remains to show that $\alpha_3 \geq \beta_3$. Distributing, we have $\alpha_3 = p(a\wedge x)\cdot p(b\wedge x)\cdot[p(a)+p(b)]$. Under our wlog assumption that $p(a\vee b) = 1$, we have $p(a)+p(b) = 1+p(a\wedge b) \geq p(a\wedge b\wedge x)+1$. Thus $\alpha_3 \geq$

$p(a\wedge x)\cdot p(b\wedge x)\cdot[p(a\wedge b\wedge x)+1] = p(a\wedge x)\cdot p(b\wedge x)\cdot p(a\wedge b\wedge x) + p(a\wedge x)\cdot p(b\wedge x) = \beta_3$ and the verification is complete.

5.5 Comments on the Improved Bound for Left∨+

It is difficult to believe that there is no simpler elementary verification of Gilio's bound, but the authors have not been able to find one. However, we can show that the bound is optimal, as follows.

Let $r,s \in (0,1)$ with $r \le s$; we want to find a probability function p and elements a,b,x in its domain such that (i) $p(x|a) = r$, (ii) $p(x|b) = s$, (iii) $p(x|a\vee b) = r\cdot s / (r+s-r\cdot s)$. We show, more specifically, that this can be done with also $x = a\wedge b$ and $p(a\vee b) = 1$.

First step: find p,a,b such that $p(a\vee b) = 1$ and (i), (ii) hold with $x = a\wedge b$, that is, (i)′ $p(a\wedge b|a) = r$, (ii)′ $p(a\wedge b|b) = s$. To do this, draw a line of unit length, extend it to the left until 1/(1+left-extension) = r, that is, the left-extension is of length $(1-r)/r$; do the same to the right until 1/(1+right-extension) = s, as in the accompanying diagram. Take S to be the line from the leftmost point to the rightmost one, a to be the initial unit line plus its left extension, b to be the same but to the right (notice that $a\wedge b$ is now the initial unit line), and measure probabilities of lines by their length relative to the total length of the extended line. Then $p(a\vee b) = 1$ and (i)′, (ii)′ hold.

Diagram for optimality of Gilio's bound for Left∨+		
Left-extension $(1-r)/r$	1	Right-extension $(1-s)/s$
a		
	b	

Second step: show that whenever $p(a\vee b) = 1$ and (i)′, (ii)′ hold then (iii) holds with $x = a\wedge b$, that is, (iii)′ $p(a\wedge b|a\vee b) = p(a\wedge b|a)\cdot p(a\wedge b|b) / [p(a\wedge b|a)+p(a\wedge b|b)-p(a\wedge b|a)\cdot p(a\wedge b|b)$. To do this, re-run the 'it-suffices-to-show' part of the verification of Gilio's bound, that is, up to display (2), but with $a\wedge b$ in place of x and = in place of ≤. It thus suffices to show that (2) holds with those modifications, that is:

$$[p(a\wedge b)\cdot p(b) + p(a\wedge b)\cdot p(a)]\cdot p(a\wedge b) = p(a\wedge b)^3+p(a\wedge b)^2.$$

But LHS $= p(a\wedge b)^2\cdot(p(a)+p(b)) = p(a\wedge b)^2\cdot(p(a\vee b)+p(a\wedge b)) = p(a\wedge b)^2\cdot(1+p(a\wedge b))$ = RHS as desired, and the verification of optimality is complete.

How does this optimal bound for Left∨+ compare with those obtained for CM and CT? We already remarked at the beginning of section 5.4 that $(r\cdot s)/(r+s-r\cdot s) > r\cdot s$ except in the limiting case that one of r, s equals 1, in which case the two sides are equal, so the bound for Left∨+ is higher than that for CT, in other words, CT is lossier than Left∨+, except in that limiting case.

On the other hand, neither Left∨+ nor CM is lossier than the other. In section 5.3 we already gave a simple example where the conclusion of CM has probability less than $r\cdot s$ and so less than the lower bound $(r\cdot s)/(r+s-r\cdot s)$ for Left∨+. It remains to give an example where the conclusion of Left∨+ has probability less than the lower bound $(r+s-1)/r$ of CM. Put $S = \{1,2,3,4,5\}$, $a = \{1,2,3,4\}$, $b = \{2,3,4,5\}$, $x = \{2,3,4\}$ and let $p(\{i\}) = 1/5$ for each $i \in S$. Then $r = p(x|a) = 3/4$, $s = p(x|b) = 3/4$ so that $(r+s-1)/r = 2/3$, while $p(x|a\vee b) = 3/5 < 2/3$ as desired.

These facts are again consonant with the qualitative ones in Hawthorne and Makinson [10] Observation 4.1, where it was shown that while Left$\vee$+ is derivable in the system **O**-plus-CT, it is not derivable in the system **O**-plus-CM, nor is CM derivable in **O**-plus-Left$\vee$+.

5.6 Summary on Improved Bounds for CM, CT, Left$\vee$+

Thus the rules CM, CT and Left$\vee$+ all have optimal bounds that do better than the sum bound, with Left$\vee$+ also doing better than CT. They are: 'divided sum' $(r+s-1)/r$ for CM due to Bourne & Parsons [5], [6]; 'product' $r{\cdot}s$ for CT going back to Adams [3]; and the more complex bound $(r{\cdot}s)/(r+s-r{\cdot}s)$ due to Gilio [7] for Left$\vee$+. The bounds for CM and Left$\vee$+ were originally established using linear algebra but can also be verified by elementary methods, as done above. In the case of Left$\vee$+, this elementary verification is quite intricate, and it would be agreeable to find a simpler one.

6 Lossiness for non-Horn rules

We now turn to some well-known non-Horn rules that are classically correct and are sometimes added to strengthen the Kraus-Lehmann system **P** of qualitative uncertain inference.

6.1 Probabilistic Soundness for Almost-Horn Rules

The rule NR of negation rationality (whenever $a \mid\!\sim x$, then either $a\wedge b \mid\!\sim x$ or $a\wedge\neg b \mid\!\sim x$) is known to be probabilistically sound and so has zero loss – see the easy verification in Hawthorne [8] or Hawthorne & Makinson [10]. But the stronger rules of disjunctive rationality DR (whenever $a\vee b \mid\!\sim x$, then either $a \mid\!\sim x$ or $b \mid\!\sim x$) and rational monotony RM (whenever $a \mid\!\sim x$, then either $a \mid\!\sim \neg b$ or $a\wedge b \mid\!\sim x$) are not probabilistically sound. How lossy are they?

Before attempting to answer this question, we need to be quite clear about the notion of probabilistic soundness for such rules. All three take the form: $a_i \mid\!\sim x_i$ (for all $i\leq n$) / $b_j \mid\!\sim y_j$ (for at least one $j\leq m$), where $n \geq 0$ and $m \geq 1$. We will call rules of that form *almost-Horn*. Thus the right side of an almost-Horn rule may be multiple, but we require that it is never empty, i.e. that $m \geq 1$. We say that an almost-Horn rule is *probabilistically sound* iff $min_{i\leq n}\{p(x_i|a_i)\} \leq max_{j\leq m}\{p(y_j|b_j)\}$ for every probability function p. In other words, iff for every such function p and every $t \in [0,1]$, if $p(x_i|a_i) \geq t$ for all $i \leq n$, then $p(y_j|b_j) \geq t$ for some $j \leq m$. Writing $a \mid\!\sim_{p,t} x$ for $p(x|a) \geq t$, this abbreviates to: if $a_i \mid\!\sim_{p,t} x_i$ for all $i \leq n$ then $b_j \mid\!\sim_{p,t} y_j$ for some $j \leq m$. In the special case that m =1, the almost-Horn rules are exactly the Horn rules, and the definition of probabilistic soundness agrees with the earlier definition. For the three non-Horn rules NR, DR and RM presented above, $n = 1$ and $m = 2$.

The above formulation may be called the *distributed* (or alternate) way of expressing an almost-Horn rule and its probabilistic soundness. As is well known, any such rule may equivalently be expressed with negative premises replacing some (but not all) of the alternates in the conclusion. For example, with suitable choices of which item in the conclusion to shift, NR becomes: whenever $a \mid\!\sim x$ and $a\wedge b \not\mid\!\sim x$, then $a\wedge\neg b \mid\!\sim x$; DR takes the form: whenever $a\vee b \mid\!\sim x$ and $a \not\mid\!\sim x$, then $b \mid\!\sim x$; and RM reads: whenever $a \mid\!\sim x$ and $a \not\mid\!\sim \neg b$, then $a\wedge b \mid\!\sim x$. This may be called the *negative* way of formulating an almost-Horn rule.

When all but one of the alternates of the conclusion has been transformed into a negative premise, we say that the negative formulation is *pointed*. Thus for NR, DR and RM, which have only two alternates in their distributed form, their negative presentations are automatically pointed; while in the general case where there may be more than two alternates, that is not the

case. For simplicity, we focus on the distributed and pointed presentations. In general, in its pointed presentation an almost-Horn rule takes the form: a_i |~ x_i (for all $i \leq n$), b_j |/~ y_j (for all $j<m, j \neq k$) / b_k |~ y_k, where $n \geq 0$ and $m \geq 1$.

The pointed presentation gives rise to a trivially equivalent definition of probabilistic soundness: if $min_{i \leq n}\{p(x_i|a_i)\} > max_{j \leq m, j \neq k}\{p(y_j|b_j)\}$ then $min_{i \leq n}\{p(x_i|a_i)\} \leq p(y_k|b_k)$. Equivalently and perhaps more intuitively: for all $t \in [0,1]$, if $min_{i \leq n}\{p(x_i|a_i)\} \geq t$ but $max_{j \leq m, j \neq k}\{p(y_j|b_j)\} < t$ then $p(y_k|b_k) \geq t$.

But, while they are equivalent, the distributed and pointed formulations give rise to quite different ways of measuring loss, as we explain in the next section.

6.2 Distributed vs Pointed Loss

We begin by illustrating the difference between distributed and pointed loss with the examples of DR and RM, and then state the difference in general form.

For DR, distributed loss is naturally measured by the drop from $p(x|a \vee b)$ to $max\{p(x|a),p(x|b)\}$, with no loss when $max\{p(x|a),p(x|b)\} \geq p(x|a \vee b)$. This is straightforward. But pointed loss is more subtle, and moreover depends on which alternate is transformed into a negative premise. Suppose that it is the alternate a |~ x that is transformed into a negative premise a |/~ x. Then a natural notion of pointed loss would consider the fall from $p(x|a \vee b)$ to $p(x|b)$ in situations where $p(x|a \vee b) > p(x|a)$. Similarly with interchanged variables when the alternate b |~ x is transformed into a negative premise b |/~ x.

For RM, distributed loss is again straightforward, measured by the drop from $p(x|a)$ to $max\{p(\neg b|a),p(x|a \wedge b)\}$. Pointed loss again depends on which alternate is transformed into a negative premise and, as the rule is not symmetric, the dependence is more important. Suppose that it is the alternate a |~ $\neg b$ that is transformed into a negative premise a |/~ $\neg b$. Then a natural definition of pointed loss would consider the fall from $p(x|a)$ to $p(x|a \wedge b)$ in situations where $p(x|a) > p(\neg b|a)$. Suppose, on the other hand that the alternate $a \wedge b$ |~ x is transformed into a negative premise $a \wedge b$ |/~ x. Then the pointed loss would consider the passage from $p(x|a)$ to $p(\neg b|a)$ in situations where $p(x|a) > p(x|a \wedge b)$.

In general, our proposed definitions for almost-Horn rules are as follows. Consider any almost-Horn rule, written in distributed form as a_i |~ x_i (for all $i \leq n$) / b_j |~ y_j (for at least one $j \leq m$), or equivalently in pointed form as a_i |~ x_i (for all $i \leq n$), b_j |/~ y_j (for all $j<m, j \neq k$) / b_k |~ y_k, where $n \geq 0$ and $m \geq 1$.

- *Distributed loss* considers the drop from $min_{i \leq n}\{p(x_i|a_i)\}$ to $max_{j \leq m}\{p(y_j|b_j)\}$. A *bound statement* for distributed loss takes the form $max_{j \leq m}\{p(y_j|b_j)\} \geq f(\{p(x_i|a_i)_{i \leq n}\})$ for some n-argument function f expressed in purely arithmetic terms.
- *Pointed loss* considers the drop from the same $min_{i \leq n}\{p(x_i|a_i)\}$ but to $p(y_k|b_k)$ in situations where $min_{i \leq n}\{p(x_i|a_i)\} > max\{p(y_j|b_j)_{j \leq m, j \neq k}\}$. A *bound statement* for it would take the form $p(y_k|b_k) \geq g(\{p(x_i|a_i)_{i \leq n}, p(y_j|b_j)_{j \leq m, j \neq k}\})$ for some $(n+m-1)$-argument function g expressed in purely arithmetic terms.

Care should be taken about conflating n-argument functions f for distributed bounds with $(n+m-1)$-argument functions g for a pointed ones. We will see shortly how this plays out for RM and DR.

Which measure of loss is more appropriate for an almost-Horn rule – distributed or pointed and, in the latter case, pointed to which component? There does not appear to be a uniform answer. On the one hand, for RM pointed loss appears to be more natural, since the rule is most naturally and commonly thought of as a weakened version of monotony with a negative premise: whenever a |~ x and a |/~ $\neg b$, then $a \wedge b$ |~ x. On the other hand, distributed loss seems

more appropriate for the rule DR, which is normally formulated in a symmetric manner with a multiple conclusion rather than a negative premise.

6.3 Pointed Bound for RM

First, we consider RM from its natural pointed perspective, that is, proceeding from $a \mid\sim x$ and $a \not\mid\sim \neg b$ to $a \wedge b \mid\sim x$. This is closely related to the Horn rule CM which, we recall, goes from $a \mid\sim x$ and $a \mid\sim b$ to $a \wedge b \mid\sim x$. Syntactically, CM can be seen as formed by taking the negative premise $a \not\mid\sim \neg b$ of RM and changing the denial into an affirmation $a \mid\sim b$ with opposite right side. From the probabilistic point of view, this amounts to replacing the requirement that $p(\neg b|a)$ is below a threshold to the requirement that $p(\neg b|a)$ at least reaches it. From the point of view of preferential consequence, it amounts to replacing an existential condition (at least one minimal a-state is a b-state) to the corresponding universal one (all minimal a-states are b-states).

We can exploit this relationship to get a pointed bound for RM directly from any bound for CM. In particular, the improved bound for CM (section 5.1) tells us that when $p(b|a) \neq 0$ then $p(x|a \wedge b) \geq [p(x|a)+p(b|a)-1]/p(b|a)$. Rewriting $p(b|a)$ as $1-p(\neg b|a)$ and simplifying a little, this becomes a pointed bound for RM, as in the display.

Pointed bound for RM
When $p(\neg b|a) \neq 1$ $p(x|a \wedge b) \geq [p(x|a)-p(\neg b|a)]/[1-p(\neg b|a)]$

The proviso, needed to ensure that the RHS is well defined, does not exclude any cases of interest to the rule RM when it is understood as pointed towards $a \wedge b \mid\sim x$. For when so understood, the only case of interest is that where $a \mid\sim x$ while $a \not\mid\sim \neg b$, which implies that $p(x|a) > p(\neg b|a)$ so in particular $p(\neg b|a) \neq 1$.

It is clear from this bound that when $p(x|a) > p(\neg b|a)$, then $p(x|a \wedge b)$ is never zero. Nevertheless, it may get arbitrarily close to zero. To see this, let $\varepsilon > 0$; we want to find p,a,b,x with $p(x|a) > p(\neg b|a)$ and $p(x|a \wedge b) < \varepsilon$. Take any positive integer n with $\varepsilon > 1/n$. Put $S = \{1,\ldots,2n\}$, put $p(i) = 1/2n$ for all $i \in S$. Take $a = \{1,\ldots,2n\}$, $b = \{1,\ldots,n\}$, $x = \{n,\ldots, 2n\}$. Then $p(x|a) = (n+1)/2n > n/2n = p(\neg b|a)$ and $p(x|a \wedge b) = 1/n < \varepsilon$ as required.

6.4 Distributed Bound for RM

RM has a very simple distributed bound, namely: $max\{p(\neg b|a), p(x|a \wedge b)\} > p(x|a)/2$ except when LHS = 0 = RHS. Unlike the bounds formulated earlier, it is a strict inequality.

Distributed bound for RM
$max\{p(\neg b|a), p(x|a \wedge b)\} > p(x|a)/2$, except when LHS = 0 = RHS

To prove this, we again apply the improved bound for CM, but with a less immediate argument. Suppose that $max\{p(\neg b|a), p(x|a \wedge b)\} \neq 0$ and that $p(\neg b|a) \leq p(x|a)/2$; we need to show $p(x|a \wedge b) > p(x|a)/2$.

We begin by disposing of several limiting cases. (1) We may assume wlog that $p(a) \neq 0$; otherwise LHS = 1 > 1/2 = RHS as needed. (2) We may assume wlog that $p(b|a) \neq 1$; otherwise, using (1), $p(\neg b|a) = 0$ and so by the first supposition, $0 \neq p(x|a\wedge b) = p(x|a) > p(x|a)/2$ as needed. (3) We have $p(b|a) \geq 0.5 > 0$; otherwise $p(\neg b|a) > 0.5$ so by the second supposition $p(x|a) > 1$. (4) Finally, $p(x|a) \neq 0$; otherwise $p(x|a)/2 = 0$ so that by the second supposition $p(\neg b|a) = 0$, contrary to (2).

We can now proceed to the main argument. Since $p(b|a) \neq 0$ as noted in (3), the improved bound for CM with variables relabelled tells us that $p(x|a\wedge b) \geq [p(x|a)+p(b|a)-1]/p(b|a)$. By the second supposition, $p(b|a) \geq 1-p(x|a)/2$, so $p(x|a)+p(b|a)-1 \geq p(x|a)/2$, so $p(x|a\wedge b) \geq p(x|a)/(2\cdot p(b|a)) > p(x|a)/2$ since $p(x|a) \neq 0$ by (4) and $p(b|a) \neq 0$ by (3) again, and we are done.

It is optimal in the following sense: for every $\varepsilon > 0$ there is a probability function p and elements a, b, x in its domain such that $max\{p(x|a\wedge b), p(\neg b|a)\} < p(x|a)/2 + \varepsilon$. To verify this, let $\varepsilon > 0$, so there is a positive integer n such that $\varepsilon > 1/n$. Take $S = \{1,2,3\}$ with $p(\{1\}) = p(\{3\}) = 1/n$ and $p(\{2\}) = (n-2)/n$, and put $a = \{1,2,3\}$, $b = \{2,3\}$, $x = \{1,3\}$. Then $p(x|a) = 2/n$ so $p(x|a)/2 = 1/n$ while $p(x|a\wedge b) = 1/n \div (n-1)/n = 1/(n-1)$ and $p(\neg b|a) = 1/n$, so $max\{p(x|a\wedge b), p(\neg b|a)\} = p(x|a\wedge b) = 1/(n-1)$. It remains to check that $1/(n-1) < 1/n + \varepsilon$, i.e. that $\varepsilon > 1/(n-1) - 1/n$. But $1/(n-1) - 1/n = n-(n-1) / n\cdot(n-1) < 1/n < \varepsilon$ by the choice of n, and we are done.

6.5 Distributed Bound for DR

As is well known, in the context of the preferential system **P** the rule RM is strictly stronger than DR, and one might be tempted to conclude that every lower bound for RM will *ipso facto* be one for DR. But the standard derivation of DR from RM, given in Lehmann and Magidor [12], uses not only RM and the probabilistically sound rules of **O**, but also Right∧+. It is not clear whether there is a derivation of DR that uses resources from **O** plus just one application of RM, which would be needed even to begin justifying such a conclusion.

Nevertheless, DR does happen to satisfy the same kind of (strict) distributed bound as we found for RM: $max\{p(x|a), p(x|b)\} > p(x|a\vee b)/2$ unless LHS = 0 = RHS..

The limiting case is immediate. To verify the principal case, assume wlog $p(x|a) \leq p(x|b)$, so $p(x|b) \neq 0$; we need to show that $p(x|a\vee b) < 2\cdot p(x|b)$. First note that $p(x|a\vee b) \leq p(x|b)\cdot[p(a|a\vee b)+p(b|a\vee b)]$ by the following chain: $p(x|a\vee b) = p((x\wedge a)\vee(x\wedge b)|a\vee b) = p(x|a)\cdot p(a|a\vee b) + p(x|b)\cdot p(b|a\vee b) - p(x\wedge a\wedge b|a\vee b) \leq p(x|b)\cdot p(a|a\vee b) + p(x|b)\cdot p(b|a\vee b) = p(x|b)\cdot[p(a|a\vee b)+p(b|a\vee b)]$. In the case that $p(a|a\vee b)+p(b|a\vee b) < 2$, this gives directly $p(x|a\vee b) < 2\cdot p(x|b)$ as desired. In the case $p(a|a\vee b)+p(b|a\vee b) = 2$ we have $p(a|a\vee b) = 1 = p(b|a\vee b)$ so $p(x|a) = p(x|a\vee b) = p(x|b)$ and again $p(x|a\vee b) < 2\cdot p(x|b)$, completing the verification.

Distributed bound for DR
$max\{p(x\|a), p(x\|b)\} > p(x\|a\vee b)/2$, except when LHS = 0 = RHS

It is optimal in the same sense: for every $\varepsilon > 0$ there is a probability function p and elements a, b, x in its domain such that $max\{p(x|a),p(x|b)\} < p(x|a\vee b)/2 + \varepsilon$. To verify this, let $\varepsilon > 0$. Then there is a positive integer n such that $\varepsilon > 1/n$. As before, take $S = \{1,2,3\}$ with $p(\{1\}) = p(\{3\}) = 1/n$ and $p(\{2\}) = (n-2)/n$, but his time put $a = \{1,2\}$, $b = \{2,3\}$, $x = \{1,3\}$. Then $p(x|a) = p(x|b) = 1/n \div (n-1)/n = 1/(n-1)$ while $p(x|a\vee b) = 2/n$ so $p(x|a\vee b)/2 = 1/n$. But $1/(n-1) < 1/n + \varepsilon$ as already checked, so we are done.

6.6 Pointed Bound for DR

Finally, we consider bounds for pointed disjunctive rationality (DR): whenever $a\vee b$ |~ x and a |/~ x, then b |~ x. Just as pointed RM may be seen as a non-Horn counterpart of the Horn rule CM, so too we can regard DR as a non-Horn counterpart of a Horn rule taking us from $a\vee b$ |~ x and a |~ $\neg x$ to b |~ x. As far as the authors are aware, this rule has not been studied, named, or even articulated in the literature; we call it *disjunctive choice* (DC). It can easily be derived from **O** plus just one application of Right∧+ so that it satisfies a sum bound. To check this we use the same subscript notation as we did in section 4 for similar verifications, and the same acronyms for the rules applied.

Suppose $a\vee b$ $|\sim_{\varepsilon 1}$ x, a $|\sim_{\varepsilon 2}$ $\neg x$; we want to get b $|\sim_{\varepsilon 1+\varepsilon 2}$ x. By LCE and RW on the second supposition we have $(a\vee b)\wedge a$ $|\sim_{\varepsilon 1}$ $\neg a\vee\neg x$. Also $(a\vee b)\wedge\neg a \vDash \neg a\vee\neg x$ so $(a\vee b)\wedge\neg a$ $|\sim_0$ $\neg a\vee\neg x$, so by XOR $a\vee b$ $|\sim_{\varepsilon 1}$ $\neg a\vee\neg x$. Hence by Right∧+ with the first supposition, $a\vee b$ $|\sim_{\varepsilon 1+\varepsilon 2}$ $(\neg a\vee\neg x)\wedge x$ so by RW $a\vee b$ $|\sim_{\varepsilon 1+\varepsilon 2}$ $\neg a\wedge x$. Clearly $(a\vee b)\wedge\neg(a\vee b)$ $|\sim_0$ $(a\vee b)$, so by WAND, $a\vee b$ $|\sim_{\varepsilon 1+\varepsilon 2}$ $(\neg a\wedge x)\wedge(a\vee b)$ and thus by RW $a\vee b$ $|\sim_{\varepsilon 1+\varepsilon 2}$ $b\wedge x$ and finally by VCM $(a\vee b)\wedge b$ $|\sim_{\varepsilon 1+\varepsilon 2}$ x so that b $|\sim_{\varepsilon 1+\varepsilon 2}$ x by LCE.

Thus $imp(x|b) \leq imp(x|a\vee b)+imp(\neg x|a)$, in other words DC satisfies the sum bound. In positive terms, $p(x|b) \geq p(x|a\vee b)+p(\neg x|a)-1$. Rewriting $p(\neg x|a)$ as $1-p(x|a)$ we thus have a simple bound for DR: $p(x|b) \geq p(x|a\vee b)-p(x|a)$.

But one can do better, obtaining the result on display by deploying Gilio's improved bound for Left∨+. It is a better bound than $p(x|b) \geq p(x|a\vee b)-p(x|a)$ because its bottom is always less than 1 except when $p(x|a) = 0$, in which limiting case the two bounds coincide.

Pointed bound for DR
$p(x\|b) \geq [p(x\|a\vee b)-p(x\|a)] / [p(x\|a)\cdot\{p(x\|a\vee b)-2\}+1]$

To derive the displayed bound, we may suppose wlog that $p(\neg x|b) \neq 0$; otherwise $p(x|b) = 1$ and the inequality holds. Take Gilio's bound for Left∨+ in its improbability version, substitute $\neg x$ for x throughout, and rewrite $imp(\neg\cdot|\cdot)$ as $p(\cdot|\cdot)$. Observing that the precondition is satisfied since $p(\neg x|b) \neq 0$, we thus have

$$p(x|a\vee b) \leq [p(x|a)+p(x|b)-2(p(x|a)\cdot p(x|b)] / [1-p(x|a)\cdot p(x|b)]$$

so that

$$p(x|a\vee b)\cdot[1-p(x|a)\cdot p(x|b)] \leq p(x|a)+p(x|b)-2\cdot p(x|a)\cdot p(x|b)$$

and thus, distributing on the left,

$$p(x|a\vee b) - p(x|a\vee b)\cdot p(x|a)\cdot p(x|b) \leq p(x|a)+p(x|b)-2\cdot p(x|a)\cdot p(x|b).$$

Rearranging:

$$p(x|a\vee b)-p(x|a) \leq p(x|b)\cdot[p(x|a)\cdot\{p(x|a\vee b)-2\}+1]$$

giving the desired:

$$p(x|b) \geq [p(x|a\vee b) - p(x|a)] / [p(x|a)\cdot\{p(x|a\vee b)-2\}+1].$$

When $p(x|a\vee b) > p(x|a)$, as must be the case if $a\vee b$ |~ x while a |/~ x in an application of DR pointed towards b |~ x, then $p(x|b) \neq 0$. However, $p(x|b) \neq 0$ may come arbitrarily close to zero, as can easily be verified. Let $\varepsilon > 0$. We want to find p,a,b,x with $p(x|a\vee b) > p(x|a)$ and $p(x|b) <$

ε. Take any positive integer n with $\varepsilon > 1/n$. Let $S = \{1,\ldots,n^2+1\}$ and put $p(i) = 1/(n^2+1)$ for all $i \in S$. Put $a = \{1,\ldots,n^2\}$, $b = \{n^2-(n-1),\ldots,n^2+1\}$, $x = \{1,\ldots, n^2-n\}\cup\{n^2+1\}$. Then $p(x|a\vee b) = (n^2-n+1)/(n^2+1) > (n^2-n)/n^2 = p(x|a)$ and $p(x|b) = 1/n < \varepsilon$ as required.

Of course, this bound on pointed DR immediately supplies one for its Horn transform DC – we need only rewrite $p(x|a)$ as $1-p(\neg x|a)$ to get $p(x|b) \geq [p(x|a\vee b)-(1-p(\neg x|a))] / [(1-p(\neg x|a))\cdot\{p(x|a\vee b)-2\}+1]$.

6.7 Other Almost-Horn rules

There are many other rules, both Horn and non-Horn, whose lossiness one may wish to determine, but they are less central than those that have been considered in this paper. We mention briefly just a few, omitting verifications.

One such Horn rule, cautious contraposition (CC), is like plain contraposition with a parameter a and an extra 'cautionary premise' $a \mid\sim \neg x$. It says: whenever $a\wedge b \mid\sim x$ and $a \mid\sim \neg x$ then $a\wedge\neg x \mid\sim \neg b$. Its pointed almost-Horn counterpart is obtained by replacing the cautionary premise $a \mid\sim \neg x$ by $a \mid\!/\sim x$: whenever $a\wedge b \mid\sim x$ and $a \mid\!/\sim x$, then $a\wedge\neg x \mid\sim \neg b$. Neither version is probabilistically sound. CC has the bound $p(\neg b|a\wedge\neg x) \geq [p(\neg x|a)+p(x|a\wedge b)-1] / p(\neg x|a)]$ with equality when $p(b|a) = 1$. This is the same as the improved bound for CM itself, with functional form $t \geq (r+s-1)/r$ where r and s are the probabilities of the premises in the right order – which for CC means $r = p(\neg x|a)$ and $s = p(x|a\wedge b)$. This bound on CC (and thus also on its non-Horn counterpart) can be obtained either by direct calculation or by analysing a derivation of CC in the system **O** plus a single application of CM.

Another rule of possible interest is cautious transitivity (CAT), which is like plain transitivity but with an added cautionary premise. It says: whenever $a \mid\sim b$, $b \mid\sim x$ and $b \mid\sim a$ then $a \mid\sim x$. This too has its pointed almost-Horn counterpart, obtained by replacing the cautionary premise $b \mid\sim a$ by $b \mid\!/\sim \neg a$. They have the bound $p(x|a) \geq p(b|a)\cdot p(x|b\wedge a) \geq p(b|a)\cdot[p(x|b)+p(a|b)-1] / p(a|b)$, which again may be obtained either by direct calculation or by analysing a derivation – this time a very short one in in system O plus one application of each of CM and CT.

A word of warning about terminology should be made here. The almost-Horn counterparts of cautious contraposition and transitivity mentioned above, which were originally articulated in Hawthorne [8], are *not* the same as those called 'rational' contraposition and transitivity in Bezzazi et al. [4]. The latter are also pointed almost-Horn rules but with quite different negative premises. They, and other rules studied in [4], are in various respects stronger than rational monotony, but are of little interest from a probabilistic point of view. For example, the rule of determinacy preservation (DP) says, in its distributive form: whenever $a \mid\sim x$, then either $a\wedge b \mid\sim x$ or $a\wedge b \mid\sim \neg x$. This has a constant function as its trivial (and optimal) distributive bound: $max\{p(x|a\wedge b), p(\neg x|a\wedge b)\} \geq 1/2$, with value 1 in the limiting case that $p(x|a) \in \{0,1\}$. In its pointed form it says: whenever $a \mid\sim x$ and $a\wedge b \mid\!/\sim x$ then $a\wedge b \mid\sim \neg x$, and has the even more trivial exact bound $p(\neg x|a\wedge b) = 1-p(x|a\wedge b)$ except in the limiting case that $p(a\wedge b) = 0$ where LHS = 1 = RHS.

6.8 Summary on Lossiness for Almost-Horn Rules

Although the distributed and pointed formulations of almost-Horn rules are equivalent as regards their probabilistic soundness, they give rise to different kinds of bounds function, which we call distributive and pointed. For the specific almost-Horn rules DR (disjunctive rationality) and RM (rational monotony) both distributed and pointed bounds may be established. The distributive bounds may be shown to be optimal for the rule, but the loss can be quite large – up to half. The pointed bounds the bounds admit the possibility of even larger loss; indeed, they

can leave the conclusion with probability arbitrarily close to zero even when the probabilities of the premises are high.

Acknowledgements

Thanks to Hanti Lin for kick-starting the investigation of optimality, and to Hans Rott for comments on a draft.

References

1. Adams, E.W.: Probability and the logic of conditionals. In: Hintikka, J.K.K. & Suppes, P. (eds.) Aspects of Inductive Logic, pp. 265-316. North-Holland, Amsterdam (1966)

2. Adams, E.W.: Four probability-preserving properties of inferences. J. Philos. Logic **25**, 1-24 (1996)

3. Adams, E.W.: A Primer of Probability Logic. Stanford: CSLI (1998)

4. Bezzazi, H., Pino Pérez, R., Makinson, D.: Beyond rational monotony: on some strong non-Horn conditions for nonmonotonic inference operations. J. Logic and Computation **7**, 605-632 (1997)

5. Bourne, R., Parsons, S.: Propagating probabilities in system P. Proceedings of the 11th International FLAIRS Conference. AAAI Org, 440-445 (1988)

6. Bourne, R., Parsons, S.: On proofs in system **P**. Internat. J. Uncertainty, Fuzziness and Knowledge-Based Systems **8**, 203-233 (2003)

7. Gilio, A.: Probabilistic reasoning under coherence in system P. Annals of Math. and Artificial Intelligence **34**, 5-34 (2002)

8. Hawthorne, J.: On the logic of nonmonotonic conditionals and conditional probabilities. J. Philos. Logic **25**, 185-218 (1996)

9. Hawthorne, J.: New Horn Rules for Probabilistic Consequence: Is O^+ Enough? In Sven Ove Hansson (ed) David Makinson on Classical Methods for Non-Classical Problems, pp. 157-166. Springer, Dordrecht (2014)

10. Hawthorne, J., Makinson, D.: The quantitative/qualitative watershed for rules of uncertain inference. Studia Logica, **86**, 249-299 (2007)

11. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**, 167-207 (1990)

12. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence **55**, 1-60 (1992)

13. Makinson, D.: Bridges from Classical to Nonmonotonic Logic. College Publications: London (2005)

14. Makinson, D.: Logical questions behind the lottery and preface paradoxes: lossy rules for uncertain inference. Synthese **186**, 511-529 (2012)

15. Paris, J., Simmonds, R.: **O** is not enough. Review of Symbolic Logic **2**, 298-309 (2009)

David Makinson (✉)
Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton St London WC2A 2AE, UK
e-mail: d.makinson@lse.ac.uk, david.makinson@gmail.com

James Hawthorne
Department of Philosophy, University of Oklahoma, 605 Dale Hall Tower, Norman, OK, 73019, USA
e-mail: hawthorne@ou.edu