THE REVIEW OF SYMBOLIC LOGIC
Volume 2, Number 2, June 2009

O IS NOT ENOUGH

J. B. PARIS and R. SIMMONDS
School of Mathematics, University of Manchester

Abstract. We examine the closure conditions of the probabilistic consequence relation of Hawthorne and Makinson, specifically the outstanding question of completeness in terms of Horn rules, of their proposed (finite) set of rules O. We show that on the contrary no such finite set of Horn rules exists, though we are able to specify an infinite set which is complete.

§1. Introduction. A consequence relation is a binary relation $\mid\!\sim$ between sentences of classical propositional logic. While there are no specific conditions that define such relations in general, the term is understood to capture some notion of inference between sentences. Many of the consequence relations discussed in the literature model forms of uncertain inference, two related and well-studied examples being the rational consequence relations and the preferential consequence relations. In this paper we consider the probabilistic consequence relations (or PCRs), which despite very different semantics show some behavior similar to the preferential and rational consequence relations. The PCRs are defined quantitatively using a probability function which can be thought of as representing an agent's belief or confidence in a sentence (as quantified by its probability), and with inferences being made on the basis of 'sufficiently high conditional belief'.

DEFINITION. *Let* SL *be the set of sentences of a finite propositional language* L. *A probability function on* SL *is a function* w*:* SL$\to$ $[0, 1]$ *such that if* $\vdash \theta$ *then* $w(\theta) = 1$, *and if* $\vdash \neg(\theta \wedge \phi)$ *then*

$$w(\theta \vee \phi) = w(\theta) + w(\phi).$$

Each probability function w *and real number* $t \in [0, 1]$ *(called the threshold in this context) generates a PCR* $\mid\!\sim_{w,t}$ *by taking*

$$\theta \mid\!\sim_{w,t} \psi \iff w(\theta \wedge \psi) \geq t\, w(\theta)$$

for all $\theta, \psi \in$ SL.

For comparison, we shall describe two other well-known families of consequence relations. The rational consequence relations are those that satisfy the Gabbay–Lehmann–Makinson properties (as formulated in Hawthorne & Makinson, 2007), where as usual $\vdash$ denotes provability in the standard Sentential Calculus:

$\theta \mid\!\sim \theta$	(reflexivity, REF)
if $\theta \mid\!\sim \phi$, $\phi \vdash \psi$ then $\theta \mid\!\sim \psi$	(right weakening, RW)
if $\theta \mid\!\sim \psi$, $\theta \dashv\vdash \phi$ then $\phi \mid\!\sim \psi$	(left classical equivalence, LCE)
if $\theta \mid\!\sim \phi \wedge \psi$ then $\theta \wedge \phi \mid\!\sim \psi$	(very cautious monotony, VCM)

Received: November 14, 2008

© 2009 Association for Symbolic Logic
doi:10.1017/S1755020309090236

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

if $\theta \mathrel{\sim\!\!\!\vdash} \psi$ and $\phi \mathrel{\sim\!\!\!\vdash} \psi$ then $\theta \vee \phi \mathrel{\sim\!\!\!\vdash} \psi$	(disjunction in premises, OR)
if $\theta \mathrel{\sim\!\!\!\vdash} \phi$ and $\theta \mathrel{\sim\!\!\!\vdash} \psi$ then $\theta \mathrel{\sim\!\!\!\vdash} \phi \wedge \psi$	(conjunction in conclusions, AND)
if $\theta \mathrel{\sim\!\!\!\vdash} \phi$ then either $\theta \mathrel{\sim\!\!\!\vdash} \neg\psi$ or $\theta \wedge \psi \mathrel{\sim\!\!\!\vdash} \phi$	(rational monotony, RM)

There is a representation theorem for the rational consequence relations in terms of ranked stoppered preferential models (Lehmann & Magidor, 1992). A preferential model is essentially a strict partially ordered set of valuations, with lower-ranked worlds being 'preferred' to higher-ranked ones. A similar representation theorem (Krauss et al., 1990) exists for the preferential consequence relations, that is, those that satisfy the family of rules P, consisting of the Gabbay–Lehmann–Makinson conditions without the rule RM. Without this rule, the models required become the stoppered preferential models, without the condition that they be ranked.

Of the rules above, REF, RW, LCE, and VCM, along with the derived rule of supraclassicality (if $\theta \vdash \phi$ then $\theta \mathrel{\sim\!\!\!\vdash} \phi$, SUP) are probabilistically sound (see Hawthorne, 2007), that is, hold for all PCRs. The rules AND, OR, and RM are not, however. Instead, Hawthorne and Makinson showed that weakened forms of the rules AND and OR do hold probabilistically, namely:

if $\theta \wedge \phi \mathrel{\sim\!\!\!\vdash} \psi$ and $\theta \wedge \neg\phi \mathrel{\sim\!\!\!\vdash} \psi$ then $\theta \mathrel{\sim\!\!\!\vdash} \psi$	(Weak OR, WOR)
if $\theta \mathrel{\sim\!\!\!\vdash} \phi$ and $\theta \wedge \neg\psi \mathrel{\sim\!\!\!\vdash} \psi$ then $\theta \mathrel{\sim\!\!\!\vdash} \phi \wedge \psi$	(Weak AND, WAND)

These two rules together with REF, RW, LCE, and VCM comprise the family O. Although O can be derived from P, and O is probabilistically sound while P is preferentially sound, this is not an indication that the PCRs are a superclass of the preferential consequence relations. This is evidenced by the rule of negation rationality which is probabilistically sound but not preferentially sound (see Hawthorne & Makinson, 2007; Makinson, 1994):

if $\theta \mathrel{\sim\!\!\!\vdash} \psi$ then $\theta \wedge \phi \mathrel{\sim\!\!\!\vdash} \psi$ or $\theta \wedge \neg\phi \mathrel{\sim\!\!\!\vdash} \psi$ (negation rationality, NR)

With the exception of RM and NR the rules in these families are finite-premised Horn rules, meaning from a finite set of conditionals we deduce a single concluding conditional. Some have classical side conditions which must also hold before the deduction is made, although when we speak of a k-premised rule we do not count the side conditions toward k. From this point on we will restrict ourselves to the finite-premised Horn rules, and by 'Horn rule' we will mean a finite Horn rule unless otherwise stated.

Hawthorne & Makinson (2007) posed the question of whether a representation or completeness theorem could be found for PCRs as had been for rational and preferential consequence relations. Their general negative answer for such a representation showed that there is no set of finite-premised Horn rules which by themselves exactly define the PCRs, and further that such a set of rules could not be complete in the sense of deriving the finite- and countably infinite-premised probabilistically sound Horn rules.

Finally however the purely finite-premised case— can every finite-premised probabilistically sound Horn rule be derived from some simpler subset of the set of all such rules?— was left open, with the suggestion that the family O might be a candidate for such a result. In the following section we provide a new condition for the soundness of a finite-premised Horn rule, and use it to show that O is not sufficient for this purpose.

§2. Soundness of Horn rules. We begin with a result concerning Horn rules that are sound for PCRs generated by particular thresholds t, and what happens to their soundness

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

as t changes. From this point on we will be working in a finite propositional language, and so we can make use of the concept of atoms of our language.[1]

DEFINITION. *An atom is a sentence of the form* $\pm p_1 \wedge \ldots \wedge \pm p_n$, *where* $p_1, \ldots, p_n$ *are the propositional variables of our language and* $\pm p_i$ *can be either* p_i *or* $\neg p_i$. *Atoms will be denoted by* α, *possibly with a subscript. Notice that*

$$w(\theta) = \sum \{w(\alpha) : \alpha \vdash \theta\},$$

where as usual we take the sum over the empty set to be zero.

THEOREM 2.1. *Let* $t \in (0, 1)$ *and suppose that we have a Horn rule that is sound for any PCR generated with threshold* t. *Then it must also be sound for any PCR generated with threshold* $1 \geq s > t$.

Proof. Take such a Horn rule and threshold t. Suppose for contradiction that there exist sentences $\theta_0, \ldots, \theta_k$ and $\phi_0, \ldots, \phi_k$ such that $\theta_0 \mid\sim \phi_0$ can be deduced from the other $\theta_i \mid\sim \phi_i$ via the Horn rule, along with a probability function w and a threshold $s > t$ such that each $\theta_i \mid\sim \phi_i$ is in the PCR $\mid\sim_{w,s}$ but $\theta_0 \mid\sim \phi_0$ is not.

We have

$$w(\theta_i \wedge \phi_i) \geq s w(\theta_i) \geq t w(\theta_i)$$

for every $i > 0$ but

$$s w(\theta_0) > w(\theta_0 \wedge \phi_0) \geq t w(\theta_0).$$

We consider two cases, when $s = 1$ and when $s < 1$.

If $s = 1$ then $w(\theta_0 \wedge \phi_0) < w(\theta_0)$, and we can pick an atom α such that $\alpha \vdash \theta_0 \wedge \neg\phi_0$ and $w(\alpha) > 0$. Let w' be the probability function which assigns probability 1 to α and 0 to all other atoms.

For each $i > 0$ we had $w(\theta_i \wedge \neg\phi_i) = 0$, and so by our choice of atom $\alpha \nvdash \theta_i \wedge \neg\phi_i$, giving $w'(\theta_i \wedge \phi_i) = w'(\theta_i) \geq t w'(\theta_i)$. But we also have $0 = w'(\theta_0 \wedge \phi_0) < t w'(\theta_0)$, and so the PCR defined by w' and threshold t contradicts our assumption.[2]

For the second case, $s < 1$, we first observe that there then exist $0 < \lambda < 1$ and $\mu > 1$ such that

$$s\lambda(1 - t) = t\mu(1 - s) \tag{1}$$

$$w(\theta_0 \wedge \phi_0)(1 - \lambda) = (\mu - 1) w(\theta_0 \wedge \neg\phi_0). \tag{2}$$

To see this notice first that since the rule holds at one threshold in $(0, 1)$ but fails at another both $w(\theta_0 \wedge \phi_0)$ and $w(\theta_0 \wedge \neg\phi_0)$ are nonzero. So for every $\lambda \in (0, 1)$ there is a corresponding μ (> 1) that satisfies equation (2).

[1] Were we to allow infinite languages and infinite sets of premises then no set of finite premise rules can provide a completeness theorem. To see this notice that if the $\alpha_i^{(n)}$ are the atoms of a propositional language with n propositional variables then $\top \mid\sim \bot$ is entailed from $\Gamma = \{ \top \mid\sim \alpha_i^{(n)} \mid 1 \leq n \in \mathbb{N},\ 1 \leq i \leq 2^n \}$ but not by any finite subset of Γ. (A similar example has already been given by, Hawthorne & Makinson, 2007, corollary 5.2.)

[2] Alternatively we could use a continuity argument in this case to show that 1 cannot be the least s for which we have such a failure.

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

We then show that at least one such pair of λ and μ must satisfy equation (1). Rearrange equation (2) to give μ as a linear function of λ and substitute this into equation (1) to obtain

$$\lambda[s(1-t)w(\theta_0 \wedge \neg\phi_0) + t(1-s)w(\theta_0 \wedge \phi_0)] = t(1-s)w(\theta_0),$$

which, since $s(1-t) > t(1-s) > 0$, has a solution in the range $\lambda \in (0, 1)$, as required.

Returning to the main proof, consider the probability function w'' which assigns probabilities to atoms in the following way:

$$w''(\alpha) = \begin{cases} \lambda w(\alpha) & \text{if} \quad \alpha \vdash \theta_0 \wedge \phi_0 \\ \mu w(\alpha) & \text{if} \quad \alpha \vdash \theta_0 \wedge \neg\phi_0 \\ w(\alpha) & \text{if} \quad \alpha \vdash \neg\theta_0 \end{cases}$$

Each $i > 0$ for which $w(\theta_i) = 0$ also has $w''(\theta_i) = 0$. For those $i > 0$ with $w(\theta_i) > 0$, the conditional probability $w''(\phi_i|\theta_i)$ is defined and equal to

$$w''(\theta_i \wedge \phi_i)/(w''(\theta_i \wedge \phi_i) + w''(\theta_i \wedge \neg\phi_i)).$$

Notice that the 'worst case' for each i—that is, the situation which would minimize this conditional probability—would occur when $\theta_i \wedge \phi_i \vdash \theta_0 \wedge \phi_0$ and $\theta_i \wedge \neg\phi_i \vdash \theta_0 \wedge \neg\phi_0$ and so we can say

$$w''(\phi_i|\theta_i) \geq \frac{\lambda w(\theta_i \wedge \phi_i)}{\lambda w(\theta_i \wedge \phi_i) + \mu w(\theta_i \wedge \neg\phi_i)}.$$

Writing equation (1) as a formula for s, substituting it into

$$w(\theta_i \wedge \phi_i)/(w(\theta_i \wedge \phi_i) + w(\theta_i \wedge \neg\phi_i)) \geq s$$

then multiplying out and simplifying gives us

$$\frac{\lambda w(\theta_i \wedge \phi_i)}{\lambda w(\theta_i \wedge \phi_i) + \mu w(\theta_i \wedge \neg\phi_i)} \geq t.$$

On the other hand, writing equation (1) as a formula for s and similarly substituting it into

$$w(\theta_0 \wedge \phi_0) < sw(\theta_0 \wedge \phi_0) + sw(\theta_0 \wedge \neg\phi_0)$$

gives

$$t > \frac{\lambda w(\theta_0 \wedge \phi_0)}{\lambda w(\theta_0 \wedge \phi_0) + \mu w(\theta_0 \wedge \neg\phi_0)} = \frac{w''(\theta_0 \wedge \phi_0)}{w''(\theta_0)},$$

and so once again the PCR defined by w'' and t contradicts our assumption. □

By its definition any PCR generated with a threshold of 0 is the 'everything relation', SL × SL. As a consequence all Horn rules are sound in such a PCR, since their conclusions are necessarily inside the relation. This leads to the following corollary of Theorem 2.1:

COROLLARY 2.2. *For a Horn rule to be probabilistically sound (for all thresholds) it need only be shown that it is sound for arbitrarily small positive thresholds.*

Our next step is to find a condition necessary and sufficient for a Horn rule to be sound for all PCRs generated with a given rational threshold. To do this we require some new notation, and a result borrowed from linear algebra.

DEFINITION. *Fix an enumeration α_j of the atoms of our language. For a sentence $\theta \in SL$ we write $\langle\theta\rangle$ for the vector in $\{0, 1\}^{2^n}$ that has jth component 1 just if $\alpha_j \vdash \theta$.*

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

For a probability function w write $\langle w\rangle$ for the corresponding vector in $[0,1]^{2^n}$ which has jth component equal to $w(\alpha_j)$. Notice that with the usual dot product $w(\theta) = \langle w\rangle.\langle\theta\rangle$.

The following result was given by Farkas (1902). Originally concerned with real vectors, we give a version here for rational vectors which may be obtained from the former by recalling that the structure $\langle \mathbb{Q}, +.0, 1, <\rangle$ is an elementary substructure of the structure $\langle \mathbb{R}, +.0, 1, <\rangle$ (see for example Marker, 2002, corollary 3.1.17).

LEMMA 2.3. Farkas' Lemma. *Given vectors $\vec{v}_1, \ldots, \vec{v}_r \in \mathbb{Q}^m$ let Q be the cone defined by $\{\sum_1^r c_i\vec{v}_i : 0 \le c_1, \ldots c_r \in \mathbb{Q}\}$. For any $\vec{g} \in \mathbb{Q}^m$ if $\vec{g} \notin Q$ then there exists $\vec{y} \in \mathbb{Q}^m$ such that $\vec{g}\cdot\vec{y} < 0$ and $\vec{h}\cdot\vec{y} \ge 0$ for any $\vec{h} \in Q$.*

PROPOSITION 2.4. *Take a Horn rule and a threshold $0 < p/q < 1$ (where $p, q \in \mathbb{N}$). The rule is sound for PCRs generated with this threshold just if for any $\theta_0 \mid\!\sim \phi_0$ that it derives from some $\theta_1 \mid\!\sim \phi_1, \ldots, \theta_k \mid\!\sim \phi_k$ there exist $N_0, \ldots, N_k \in \mathbb{N}$ with $N_0 > 0$ such that*

$$qN_0\,\langle\theta_0 \wedge \phi_0\rangle - pN_0\,\langle\theta_0\rangle \ge \sum_{i=1}^{k} N_i\,(q\langle\theta_i \wedge \phi_i\rangle - p\langle\theta_i\rangle)\,. \tag{3}$$

(Here vector inequalities such as this should be interpreted line-by-line, i.e., as a system of 2^n standard inequalities.)

Proof. Take such premises and conclusion. In vector notation, this deduction is sound at threshold p/q just if for any probability function w, if whenever

$$\langle\theta_i \wedge \phi_i\rangle.\langle w\rangle \ge \frac{p}{q}\ \langle\theta_i\rangle.\langle w\rangle \quad \text{for } 1 = 1, \ldots, k$$

then

$$\langle\theta_0 \wedge \phi_0\rangle.\langle w\rangle \ge \frac{p}{q}\ \langle\theta_0\rangle.\langle w\rangle.$$

Equivalently, for all probability functions w, if whenever

$$(q\langle\theta_i \wedge \phi_i\rangle - p\langle\theta_i\rangle).\langle w\rangle \ge 0 \quad \text{for } 1 = 1, \ldots, k$$

then

$$(q\langle\theta_0 \wedge \phi_0\rangle - p\langle\theta_0\rangle).\langle w\rangle \ge 0.$$

We can show that this soundness statement is equivalent to the following condition:

Let $e_1, \ldots, e_{2^n}$ be the standard basis vectors for $\mathbb{Q}^{2^n}$. Let Q be the cone defined as in Lemma 2.3 by these basis vectors and by $q\langle\theta_i \wedge \phi_i\rangle - p\langle\theta_i\rangle$ for each $i > 0$. Our condition is then that $q\langle\theta_0 \wedge \phi_0\rangle - p\langle\theta_0\rangle$ is inside Q.

To see this, first suppose that this vector was not in Q. Consider the $\vec{y}$ that Lemma 2.3 shows to exist for this vector. Since $\forall j\ \vec{y}.\vec{e}_j \ge 0$ it has all nonnegative components, and by its other properties some of these must be positive. Then the soundness statement fails for the probability function w given by $\langle w\rangle = \vec{y}/\sum_i y_i$.

For the converse, suppose the vector is inside Q. Take any w for which

$$(q\langle\theta_i \wedge \phi_i\rangle - p\langle\theta_i\rangle).\langle w\rangle \ge 0 \quad \text{for} \quad 1 = 1, \ldots, k.$$

Then the scalar product of $\langle w\rangle$ with any of the vectors defining the cone is nonnegative, and so its scalar product with any vector inside the cone is nonnegative, meaning

$$(q\langle\theta_0 \wedge \phi_0\rangle - p\langle\theta_0\rangle).\langle w\rangle \ge 0.$$

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

The vector being inside the cone Q is then in turn equivalent to: $\exists c_i \in \mathbb{Q}$ for each $i = 1, \ldots, k$ such that $c_i \geq 0$ and

$$q\langle\theta_0 \wedge \phi_0\rangle - p\langle\theta_0\rangle \geq \sum_{i=1}^{k} c_i q\langle\theta_i \wedge \phi_i\rangle - c_i p\langle\theta_i\rangle,$$

since we can bring this statement to equality by adding appropriate terms of the form $d_j \vec{e}_j$ (some $0 \leq d_j \in \mathbb{Q}$) to the right hand side, which is then just the definition of inclusion of the vector in Q.

Considering the c_i as irreducible fractions, we multiply by each of their denominators to get $N_0, \ldots, N_k \in \mathbb{N}$ such that $N_0 > 0$ and

$$qN_0\langle\theta_0 \wedge \phi_0\rangle - pN_0\langle\theta_0\rangle \geq \sum_{i=1}^{k} qN_i\langle\theta_i \wedge \phi_i\rangle - pN_i\langle\theta_i\rangle,$$

as required. □

We can now combine Corollary 2.2 and Proposition 2.4 to obtain a condition that determines whether or not a Horn rule is probabilistically sound for all thresholds. To do this it is neatest to move to a nonstandard elementary extension M of the structure comprising $\mathbb{N}, \mathbb{Q}, \mathbb{R}, +, \times, <, 0, 1$ etc., such as one might obtain by taking a suitable ultrapower of this structure. Let q be a nonstandard natural number in $\mathbb{N}^M$ (= the set of natural numbers according to M) so $0 < 1/q < t$ for all $0 < t \in \mathbb{R}$.

Then by a standard overspill argument (see Pillay, 1981, p. 10) and using Corollary 2.2 and Proposition 2.4 we have that

THEOREM 2.5. *The following are equivalent:*

(a) For all $0 < t \leq 1$ *and probability functions* w *if* $\theta_1 \mid\!\sim_{w,t} \phi_1, \ldots, \theta_k \mid\!\sim_{w,t} \phi_k$ *then* $\theta_0 \mid\!\sim_{w,t} \phi_0$.

(b) There are $N_0, N_1, \ldots, N_k \in \mathbb{N}^M$ *such that (in* M*)*

$$qN_0\langle\theta_0 \wedge \phi_0\rangle - N_0\langle\theta_0\rangle \geq \sum_{i=1}^{k} N_i\left(q\langle\theta_i \wedge \phi_i\rangle - \langle\theta_i\rangle\right). \tag{4}$$

§3. A completeness result. The plan now is to rephrase this part (b) of Theorem 2.5 into a friendlier form and ultimately return to the standard universe.

With the notation as above consider the following table:

$\theta_0 \wedge \phi_0$	$\theta_1 \wedge \neg\phi_1$	$\ldots$	$\theta_k \wedge \neg\phi_k$
N_0	$(q-1)N_1$	$\ldots$	$(q-1)N_k$
$\theta_0 \wedge \neg\phi_0$	$\theta_1 \wedge \phi_1$	$\ldots$	$\theta_k \wedge \phi_k$
$N_0(q+1)$	$(2q-1)N_1$	$\ldots$	$(2q-1)N_k$
$\neg\theta_0$	$\neg\theta_1$	$\ldots$	$\neg\theta_k$
$N_0 q$	$N_1 q$	$\ldots$	$N_k q$

each cell of which contains a sentence and an associated 'cost' in terms of q and the N_i. Define a 'path' to be a choice of one cell from each column of the table, and the cost of a path to be the sum of its cells' costs.

Condition ($\star$) is that for any path whose cost exceeds $q\sum_{i=0}^{k} N_i$ the sentences on the path must form a classically inconsistent set.

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

PROPOSITION 3.1. *Condition (⋆) and (4) are equivalent.*

Proof. Notice that each column of the table contains three disjoint sentences whose disjunction is a tautology, and so every atom classically implies exactly one sentence from each column. Condition (⋆) is therefore equivalent to saying that for all atoms α, the path of sentences that are implied by α has cost at most $q\sum_{=0}^{k} N_i$.

For a given atom, let $X, Y, Z \subseteq \{1, \ldots, k\}$ be the sets of nonzero column indices for which the atom implies the top, middle, and bottom sentence in the table respectively. Condition (⋆) is then equivalent to: For all atoms

$$sN_0 - \sum_X (q-1)N_i + \sum_Y (2q-1)N_i + \sum_Z qN_i \leq q\sum_{i=0}^{k} N_i \tag{5}$$

where $s = 1, q+1$ or q depending which zeroth-column sentence the atom implies.

For an atom α_j, notice that $s = 1$ just if $\langle\theta_0\rangle_j = 1$ and $\langle\neg(\theta_0 \wedge \phi_0)\rangle_j = 0$. Similarly, $s = q+1$ just if both these jth coordinates are 1, and $s = q$ just if $\langle\theta_0\rangle_j = 0$ and $\langle\neg(\theta_0 \wedge \phi_0)\rangle_j = 1$.

In the same way, a column index $i > 0$ is in X just if $\langle\theta_i\rangle_j = 1$ and $\langle\neg(\theta_i \wedge \neg\phi_i)\rangle_j = 0$, is in Y just if both are 1, and is in Z just if $\langle\theta_i\rangle_j = 0$ and $\langle\neg\theta_i \wedge \neg\phi_i\rangle_j = 1$.

As such, the system of 2^n inequalities (one for each atom) described by (5)—which was equivalent to condition (⋆)—is in turn equivalent to the vector inequality

$$N_0\langle\theta_0\rangle + qN_0\langle\neg(\theta_0 \wedge \phi_0)\rangle + \sum_{i=1}^{k}(q-1)N_i\langle\theta_i\rangle + \sum_{i=1}^{k} qN_i\langle\neg(\theta_i \wedge \neg\phi_i)\rangle \leq \sum_{i=0}^{k} qN_i\langle\top\rangle.$$

Finally, by subtracting

$$N_0\langle\theta_0\rangle + qN_0\langle\neg(\theta_0 \wedge \phi_0)\rangle + \sum_{i=1}^{k} qN_i\langle\top\rangle$$

from each side and simplifying, we get the inequality (4) as another equivalent condition. □

The equivalence proved in Proposition 3.1 immediately gives us a sound and complete set of rules, namely for each choice of $0 < k \in \mathbb{N}$ and $N_0, N_1, \ldots, N_k \in \mathbb{N}^M$ the $(N_0, N_1, \ldots, N_k)$ rule which asserts that:

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_k \mid\sim \phi_k}{\theta_0 \mid\sim \phi_0}$$

whenever the classical side condition (⋆) holds.

Unfortunately for practical purposes this seems of limited use since we do not have any limits on the values (standard or nonstandard) that might be required for these N_0, $N_1, \ldots, N_k$. However we now show that we can take N_i to be standard degree polynomials in q with coefficients in the integers, $\mathbb{Z}$.

To see this suppose that the required side conditions hold for the $(N_0, N_1, \ldots, N_k)$ rule to give $\theta_0 \mid\sim \phi_0$ from $\theta_1 \mid\sim \phi_1, \ldots, \theta_k \mid\sim \phi_k$. Now list as

$$\sum_{i\in S} a_i(q)X_i \geqslant 0 \tag{6}$$

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

all the inequalities in variables $X_0, X_1, \ldots, X_k$ satisfied by $N_0, N_1, \ldots, N_k$ and appearing as path costs for this rule (where $\geqslant$ is either $>$ or $\geq$) along with $X_0 > 0$ and $X_i \geq 0$ for all $i > 0$, so $a_i(q)$ here is one of $1, q, q-1, q+1, 2q-1$.

Starting with X_k we rewrite and divide the subset of our inequalities involving X_k as

$$X_k \geqslant \sum_{i \in S,\ i \neq k} \frac{a_i(q)}{a_k(q)} X_i \tag{7}$$

when $a_k(q) > 0$ and

$$\sum_{i \in S\ i \neq k} \frac{a_i(q)}{a_m(q)} X_i \geqslant X_k \tag{8}$$

when $a_k(q) < 0$. Now suppose that out of these inequalities the right hand side expression in (7) which is overall the largest when evaluated at $N_0, N_1, \ldots, N_{k-1}$ is

$$\sum_{i \in S',\ i \neq k} \frac{a'_i(q)}{a'_k(q)} X_i,$$

which exists, since we have $X_k \geq 0$ as one of our inequalities in (7), and is nonnegative. Similarly let

$$\sum_{i \in S''\ i \neq k} \frac{a''_i(q)}{a''_k(q)} X_i$$

be the left hand side expression in (8) which is smallest when evaluated at $N_0, N_1, \ldots, N_{k-1}$. (If there is no inequality of type (8), take it to be

$$\sum_{i \in S',\ i \neq k} \frac{a'_i(q)+1}{a'_k(q)} X_i$$

where the a'_i etc. arise from the previous choice of 'largest' and this did not come from $X_k > 0$, otherwise take it to be just X_0.)

Then we see that if

$$J_k = 2^{-1}\left(\left(\sum_{i \in S',\ i \neq k} \frac{a'_i(q)}{a'_k(q)} N_i\right) + \left(\sum_{i \in S''\ i \neq k} \frac{a''_i(q)}{a''_k(q)} N_i\right)\right)$$

then exactly the same inequalities in (6) hold for $N_0, N_1, \ldots, N_{k-1}, J_k$ as did for $N_0, N_1, \ldots, N_{k-1}, N_k$. In particular then the $(N_0, N_1, \ldots, N_{k-1}, J_k)$ rule gives $\theta_0 \mid\sim \phi_0$ from $\theta_1 \mid\sim \phi_1, \ldots, \theta_k \mid\sim \phi_k$.

Now replace X_k in (6) by

$$2^{-1}\left(\left(\sum_{i \in S',\ i \neq k} \frac{a'_i(q)}{a'_k(q)} X_i\right) + \left(\sum_{i \in S''\ i \neq k} \frac{a''_i(q)}{a''_k(q)} X_i\right)\right).$$

This gives a set of inequalities similar to those in (6), except that the coefficients now come from a finite choice of rational functions with integer coefficients (in M), and satisfied by $N_0, N_1, \ldots, N_{k-1}, J_k$.

Now iterate this process successively removing the variables $X_{k-1}, X_{k-2}, \ldots, X_2, X_1$ until we are left with just X_0. At that stage we will have a solution to (6) in which, after substituting back, all the X_i are positive and of the form $f(q)X_0$ where the $f(q)$ come

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

from some finite set of rational functions in q with coefficients from the standard integers $\mathbb{Z}$. Setting X_0 to be the product of the (positive) denominators of these rational functions now gives a solution to (6) as polynomials $F_i(q)$ over $\mathbb{Z}$ in q and so necessarily $\theta_0 \mid\sim \phi_0$ follows from $\theta_1 \mid\sim \phi_1, \ldots, \theta_k \mid\sim \phi_k$ by the $(F_0(q), F_1(q), \ldots, F_k(q))$ rule.

Clearly we can put a bound (depending on k) on the degrees of these polynomials $F_i(q)$ and the sizes of their coefficients. In this sense then we can claim to have an at least 'theoretically graspable' complete set of Horn rules though with our current best estimate of these bounds (double exponential) it can scarcely be said to be 'practically graspable'!

Fortunately there is a trick for producing an apparently simpler complete set of rules. Namely, notice that if we apply the $(N_0, N_1, N_2, N_3, \ldots, N_k)$ rule with $\theta_1 \mid\sim \phi_1$ and $\theta_2 \mid\sim \phi_2$ actually the same premise then this is equivalent to applying the $(N_0, N_1 + N_2, N_3, \ldots, N_k)$ rule with $\theta_1 \mid\sim \phi_1$ the premise corresponding to $N_1 + N_2$. Similarly an application of the $(N_0, N_1, N_2, \ldots, N_k)$ rule with $\theta_1 \mid\sim \phi_1$ being $\bot \mid\sim \bot$ (which follows from the (1) rule) is equivalent to using the $(N_0, N_2, \ldots, N_k)$ rule without this premise.

Given these observations we see that a complete set of rules arises just from rules of the form $(N_0, N_1, \ldots, N_k)$ where for some $m \in \mathbb{N}$ N_0 is a polynomial in q of degree at most m, positive lead coefficient, and integer coefficients with modulus at most m and $N_1, N_2, \ldots, N_k$ consist of m copies of $q^i(q - m)$ for each $0 \leq i < m$ and m^2 copies of q^i for each $0 \leq i \leq m$, the reason being that by choosing m large enough we can combine these building blocks to produce any finite family of polynomials in q over $\mathbb{Z}$ with positive lead coefficients.

Despite the evident improvement here, in terms of the number of necessary k premise rules, on what we had initially we obviously still have an awfully complicated set of rules. Devices exist for showing many of these to be reducible to simpler rules (e.g., fewer premises or lower degree polynomials) though we have no general result indicating that a significantly more manageable, or comprehensible, set exists. For specific small k however this can be possible, as we now indicate in the case $k < 3$, a relevant case here since the rules of O all fall under that mantle.

§4. Horn rules with fewer than three premises.

THEOREM 4.1. *All probabilistically sound Horn rules with two or fewer premises are derivable in one step using the* (1)*,* (1, 1)*,* (1, 1, 1) *and* (1, 1, q) *rules.*

Proof. Suppose that it is probabilistically sound to conclude $\theta_0 \mid\sim \phi_0$ from $\theta_1 \mid\sim \phi_1$ and $\theta_2 \mid\sim \phi_2$. Then as above this must be an instance of some (N_0, N_1, N_2) rule with $N_0, N_1, N_2 \in \mathbb{N}^M$, where by reordering the premises we may assume that $N_2 \geq N_1$. We now consider various cases.

If (in M), $N_0 > N_1 + N_2$ then the conditions in the (N_0, N_1, N_2) rule imply that $\theta_0 \vdash \phi_0$, in which case $\theta_0 \mid\sim \phi_0$ follows immediately from the empty set of premises by the (1) rule. So assume now that $N_0 \leq N_1 + N_2$.

If $N_0 > N_1$ then the conditions will be fulfilled for $\theta_0 \mid\sim \phi_0$ to follow immediately from $\theta_2 \mid\sim \phi_2$ by the (1, 1) rule. So assume further that $N_0 \leq N_1 \leq N_2$.

If now it is not the case that $\theta_0 \mid\sim \phi_0$ follows directly from $\theta_1 \mid\sim \phi_1$ and $\theta_2 \mid\sim \phi_2$ by the (1, 1, 1) rule it must be because $N_1(q - 1) \leq N_2$. But then the conditions in the (N_0, N_1, N_2) rule imply that instead $\theta_0 \mid\sim \phi_0$ follows immediately from $\theta_1 \mid\sim \phi_1$ and $\theta_2 \mid\sim \phi_2$ by the (1, 1, q) rule. □

For example, the right-weakening rule RW can be derived in one step from the (1, 1) rule. Examining condition ($\star$) of Proposition 3.1 for this tuple it can be seen that a

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

conclusion $\theta_0 \mid\sim \phi_0$ follows a premise $\theta_1 \mid\sim \phi_1$ via this rule iff $\theta_0 \wedge \neg\phi_0 \vdash \theta_1$ and $\theta_1 \wedge \phi_1 \vdash \theta_0 \wedge \phi_0$. Since this must hold whenever $\theta_0 = \theta_1$ and $\phi_1 \vdash \phi_0$, that is, whenever the rule RW applies, we have our derivation in one step.

A similar analysis is possible with three premise rules and so on though it becomes progressively messier and (apparently) without pointing to any obvious pattern. Clearly however on this evidence the bound established earlier for the necessary k premise rules appears wildly excessive.

A natural question now is to ask how the above-mentioned (1), (1, 1), (1, 1, 1), and (1, 1, q) rules relate to O. Clearly Proposition 4.1 shows that the rules of O can be replaced by these rules in any derivation, indeed it is easy to see that RW, LCE, and VCM are weakenings of the (1, 1) rule and REF, WOR, and WAND are weakenings of the (1), (1, 1, 1), and (1, 1, q) rules respectively.

In fact a converse of this result exists, in that each of the (1), (1, 1), (1, 1, 1), and (1, 1, q) rules can be derived (in several steps) from the rules of O, see Simmonds (2009). Therefore we can observe as a corollary to Proposition 4.1 that the rules of O are 'complete enough' to derive all sound Horn rules with fewer than three premises.

Returning to the question of a simple and meaningful description of the rules required to ensure completeness it might have been hoped that at least some *finite* set would suffice. However in the next section we show that this is not the case.

§5. Nonfinitizability. We shall demonstrate the nonexistence of a finite set of complete Horn rules via a string of lemmas.

LEMMA 5.1. *Suppose that* $\theta_2 \mid\sim \phi_2$ *follows from* $\theta_1 \mid\sim \phi_1$ *and that* $\theta_3 \mid\sim \phi_3$ *follows from* $\theta_2 \mid\sim \phi_2$*, both derivations being a single use of the* (1, 1) *rule. Then*

(a) $\theta_1 \wedge \phi_1 \vdash \theta_2 \wedge \phi_2$.
(b) $\theta_2 \wedge \neg\phi_2 \vdash \theta_1 \wedge \neg\phi_1$.
(c) $\neg\theta_1 \vdash \neg\theta_2 \vee (\theta_2 \wedge \phi_2)$.
(d) $\theta_3 \mid\sim \phi_3$ *follows from* $\theta_1 \mid\sim \phi_1$ *by the* (1, 1) *rule.*

Proof. Because the (1, 1) rule immediately gives $\theta_2 \mid\sim \phi_2$ from $\theta_1 \mid\sim \phi_1$ the threshold conditions give that each of the pairs

$$\{\theta_2 \wedge \neg\phi_2, \theta_1 \wedge \phi_1\}, \quad \{\theta_2 \wedge \neg\phi_2, \neg\theta_1\}, \quad \{\neg\theta_2, \theta_1 \wedge \phi_1\}$$

are inconsistent. Between them these give (a), (b), (c) and using these twice the required threshold conditions for (d) follow. □

LEMMA 5.2. *Suppose that*

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_r \mid\sim \phi_r}{\theta_0 \mid\sim \phi_0} \quad \textit{and} \quad \frac{\psi \mid\sim \eta}{\theta_r \mid\sim \phi_r} \tag{9}$$

are an instance of the $(m_0, m_1, \ldots, m_r)$ *rule and* (1, 1) *rule respectively. Then*

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_{r-1} \mid\sim \phi_{r-1}, \psi \mid\sim \eta}{\theta_0 \mid\sim \phi_0} \tag{10}$$

is also an instance of the $(m_0, m_1, \ldots, m_r)$ *rule.*

Proof. Consider a path through the diagram for (10) which exceeds the threshold for this rule. If this path 'goes through' $\psi \wedge \neg\eta$ then even ignoring this contribution the remaining

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

part of the path must exceed the threshold and so be inconsistent since it already appears in (9). On the other hand if this path goes through $\psi \wedge \eta$ or $\neg\psi$ then conditions (a) and (c) in Lemma 5.1 above, and the fact that this(ese) corresponding path(s) are contradictory in (9), ensure that the path remains contradictory. □

LEMMA 5.3. *Suppose that*

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_r \mid\sim \phi_r}{\theta_0 \mid\sim \phi_0} \tag{11}$$

is an instance of the $(m_0, m_1, \ldots, m_r)$ *rule and* $\theta_r \vdash \phi_r$. *Then*

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_{r-1} \mid\sim \phi_{r-1}}{\theta_0 \mid\sim \phi_0} \tag{12}$$

is an instance of the $(m_0, m_1, \ldots, m_{r-1})$ *rule.*

Proof. Consider a path through (12) which exceeds its threshold. The extension of that path in (11) through $\psi \wedge \neg\eta$ is certainly contradictory, since $\psi \wedge \neg\eta$ is, and the other two possible extensions are also contradictory because they exceed the threshold for (11). Hence the original path through (12) must be contradictory, as required. □

THEOREM 5.4. *For* $k \geq 2$ *the* $(k, 1, 1, \ldots, 1)$ *rule* ($k+1$ 1*'s) is not derivable from rules with fewer premises.*

Proof. Take an instance

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_k \mid\sim \phi_k}{\theta_0 \mid\sim \phi_0} \tag{13}$$

of the $(k, 1, 1, \ldots, 1)$ rule where the only inconsistent paths are those that exceed the threshold for this rule (according to condition $\star$). If, say,

$$\frac{\theta_1 \mid\sim \phi_1, \ldots, \theta_r \mid\sim \phi_r}{\psi \mid\sim \eta} \tag{14}$$

is an instance of a rule with fewer premises using a subset of the premises from (13) then it must be the case that either $\psi \vdash \eta$ or $\psi \mid\sim \eta$ follows from just some one $\theta_i \mid\sim \phi_i$ of these premises by the (1, 1) rule. But then as Lemmas 5.2 and 5.3 show we can always remove $\psi \mid\sim \eta$ or replace it by $\theta_i \mid\sim \phi_i$ in any subsequent use of it as a premise in its own right.

The result follows since for $k \geq 2$ the chosen instance of the $(k, 1, 1, \ldots, 1)$ rule the conclusion does not follow from just one, or none, of the premises using just the $\langle 1, 1\rangle$ or $\langle 1\rangle$ rules. □

For example, the deduction from

$$(p \wedge \neg q) \vee (\neg p \wedge q) \mid\sim r, \quad p \mid\sim r, \quad q \mid\sim r \quad \text{to} \quad p \vee q \mid\sim r,$$

where p, q, and r are propositional variables, is sound by the (2, 1, 1, 1) rule. However, no paths with cost less than the threshold for this rule are inconsistent for these formulae. As such then it cannot be derived using only Horn rules with fewer than three premises.

§6. Conclusion. Given the naturalness of the the consequence relation $\mid\sim$ it is perhaps rather surprising that any complete set of Horn rules must be infinite and, as far as our current knowledge goes, somewhat incomprehensible. Of course it may be that further

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press

research will provide an easy-to-grasp schematic set of such rules, and maybe that the use of non-Horn rules will enable us to finitize the system.

One such rule that has been suggested by Hawthorne and Makinson is the negation rationality rule, NR, given earlier. However, whether or not adopting this rule enables us to remove all but a finite number of $(N_0, N_1, \ldots, N_k)$ rules remains an intriguing open question. The fact that up to now we have not been able to derive the $(k, 1, 1, \ldots, 1)$ rules $(k \geq 2)$ mentioned earlier from from O + NR suggests a negative answer here.

§7. Acknowledgments. We would like to thank David Makinson and Jim Hawthorne for their unflagging encouragement and advice. Supported by a UK Engineering and Physical Sciences Research Council Research Studentship.

BIBLIOGRAPHY

Farkas, J. (1902). Theorie der Einfachen Ungleichungen. *Journal für die Reine und Angewandte Mathematik*, **124**, 1–27.

Hawthorne, J. (1996). On the logic on non-monotonic conditionals and conditional probabilies. *Journal of Philosophical Logic*, **25**, 185–218.

Hawthorne, J. (2007). Nonmonotonic conditionals that behave like conditional probabilities above a threshold. *Journal of Applied Logic*, **5**, 625–637.

Hawthorne, J., & Makinson, D. (2007). The qualitative/quantitative watershed for rules of uncertain inference. *Studia Logica*, **86**, 247–297.

Krauss, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, **44**, 167–207.

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

Makinson, D. (1994). General patterns in nonmontonic reasoning. In Gabbay, Hogger, and Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 3. Oxford: Clarendon Press, pp. 35–110.

Marker, D. (2002). *Model Theory: An Introduction*. Graduate Texts in Mathematics, Vol. 217. New York: Springer-Verlag.

Pillay, A. (1981). Models of Peano Arithmetic, In editors Berline, McAloon, and Ressayre, editors. *Model Theory and Arithmetic*, Lecture Notes in Mathematics, Vol. 890. Berlin: Springer-Verlag, pp. 1–12.

Simmonds, R. (2009). On Horn Closure Conditions for Probabilistic Consequence Relations, University of Manchester. Available from: http://www.maths.manchester.ac.uk/~jeff/

SCHOOL OF MATHEMATICS
UNIVERSITY OF MANCHESTER
MANCHESTER M13 9PL, UK
E-mail: jeff.paris@manchester.ac.uk

SCHOOL OF MATHEMATICS
UNIVERSITY OF MANCHESTER
MANCHESTER M13 9PL, UK
E-mail: richard.simmonds@gmail.com

https://doi.org/10.1017/S1755020309090236 Published online by Cambridge University Press