When maximizing entropy gives the rational closure

L.C.Hill* and J.B.Paris
Department of Mathematics
University of Manchester
Manchester M13 9PL
UK
Email lhill@ma.man.ac.uk

May 17, 2006

Abstract

We introduce a generalization of ϵ-probability functions and a new correspondence with rational consequence relations. The rational consequence relations corresponding to the generalized ϵ-probability functions of maximum entropy are investigated and it is shown that in this case the analogous notion of 'maximum entropy closure' equals the rational closure, thus reconfirming the rational closure of a conditional knowledge base as the simplest, least prejudiced, rational consequence relation satisfying that knowledge base.

Introduction

In their seminal publications Lehmann and Magidor, [8], and Pearl, [10], proffered the *rational closure* of a conditional knowledge base K as the simplest, least informative, rational consequence relation satisfying K (formal definitions of these technical terms will follow shortly). Putting

*Supported by EPSRC Research Studentship 99801507.

it another way, given that K summed up all that an agent knew and was willing a priori to assume and that this agent's notion of 'consequence' was to be a rational consequence relation then this relation *should* be the rational closure of K.

Despite its naturalness the claim of the rational closure to fulfill this role has been brought into question because of its failure to satisfy property inheritance to exceptional subclasses. For example, it was argued that knowing (only) that Swedes are usually tall and fair haired should justify the conclusion that short Swedes are also usually fair. That is, that the exceptional subclass of short Swedes should still inherit the common feature of Swedes of fairness of hair.

In order to redress this apparent deficiency various other 'closures' were proposed amongst which was the *maximum entropy closure* of Goldzsmidt, Morris and Pearl. This maximum entropy closure was based on an alternate derivation of rational consequence relations (ϵ-semantics) introduced by Pearl, see [10] (see also the appendix to [8]) which associated rational consequence relations with so called ϵ-probability functions. As the name suggests the maximum entropy closure was defined as that rational consequence relation satisfying K with (an) associated ϵ-probability function of largest possible entropy. As Goldzsmidt et al showed in [5] in the case of the tall fair Swedes the maximum entropy closure does successfully condone the conclusion that even short Swedes are normally fair. Furthermore, in view of the work of Shannon [11], [12], identifying entropy with uncertainty, or lack of information, this closure would seem to be entirely justified. In short, it is the rational consequence relation satisfying K whose information content is as small as possible.

Despite this apparent triumph however we were left with some irritating qualms. Firstly, it is not clear that if one could put oneself in the position that *all one knew* was that Swedes are normally tall and fair that one really would wish to conclude that even short Swedes are normally fair. After all, if one knew only that birds normally weigh less than 20 kilos and fly would one really wish to conclude that even birds weighing at least 20 kilos normally fly? So this argument *per se* for the desirability of inheritance to exceptional subclasses seems to us simply unjustified[1].

[1]That is not to say of course that there may not be alternate, and sounder, arguments.

Secondly the common conception of 'the maximum entropy solution' is of a solution which is minimally prejudiced, which goes as little as possible beyond what the constraints already give. Yet here, in the Swedes example, the maximum entropy solution is introducing a 'prejudice' (that even short Swedes are normally fair) which we already know from the example of the rational closure is not forced by the constraints. Indeed pushing this argument a little harder should not the maximum entropy solution really be the same thing as the rational closure?

In this paper we shall consider this question, showing, amongst other results, that the notion of 'the maximum entropy solution' can, more correctly in our opinion, be formulated in such a way that this dichotomy disappears. Before doing that however we need to introduce and recall some notation, definitions and background.

Background and notation

Throughout we shall work in a fixed finite propositional language L, with propositional variables $p_1, p_2, ..., p_n$ say, denoting by SL the set of sentences of L. A *conditional*, or *default*, is an expression of the form $\theta \mid\!\sim \phi$, where $\theta, \phi \in SL$, whose intended interpretation is that 'if θ then usually, or normally, ϕ'.

It has been argued (see, for example, [8]) that the relationship appropriate to this interpretation is that of *rational consequence relation.* Precisely, a binary relation $\mid\!\sim$ on $SL \times SL$ is a *rational consequence relation* if $\mid\!\sim$ satisfies the Gabbay-Makinson Rules (GM):-

REF $\quad \theta \mid\!\sim \theta$

LLE $\quad \dfrac{\theta \mid\!\sim \varphi \qquad \theta \equiv \psi}{\psi \mid\!\sim \varphi}$ $\qquad RWE$ $\quad \dfrac{\theta \mid\!\sim \varphi \qquad \varphi \models \psi}{\theta \mid\!\sim \psi}$

AND $\quad \dfrac{\theta \mid\!\sim \varphi \qquad \theta \mid\!\sim \psi}{\theta \mid\!\sim \varphi \wedge \psi}$ $\qquad OR$ $\quad \dfrac{\theta \mid\!\sim \varphi \qquad \psi \mid\!\sim \varphi}{\theta \vee \psi \mid\!\sim \varphi}$

CMO $\quad \dfrac{\theta \mid\!\sim \varphi \qquad \theta \mid\!\sim \psi}{\theta \wedge \varphi \mid\!\sim \psi}$ $\qquad RMO$ $\quad \dfrac{\theta \mid\!\sim \varphi \qquad \theta \not\mid\!\sim \neg\psi}{\theta \wedge \psi \mid\!\sim \varphi}$

One particular rational consequence relation we shall need to refer to later is the *trivial rational consequence relation*, $\mid\!\sim_0$, which holds between *all* θ and ϕ.

Throughout we shall use K to denote a finite set of conditionals, $\theta \mid\!\sim \phi$, where $\theta, \phi \in SL$. Notice that K is always satisfied by *some* rational consequence relation, namely $\mid\!\sim_0$. In the case that this is the only rational consequence relation satisfying K we have nothing more to say, so henceforth assume that K is consistent in the sense that it is satisfied by some non-trivial rational consequence relation.

Numerous further characterizations of rational consequence relations have been formulated, but for this paper we need only to recall the original approaches due to Kraus, Lehmann and Magidor, see [7], [8], and Pearl, [10].

Starting with the Lehmann et al approach, we shall call a sentence of L of the form

$$\pm p_1 \wedge \pm p_2 \wedge ... \wedge \pm p_n$$

an *atom* of L. Atoms of L will be denoted by α and the set of atoms of L by At^L. In the obvious way an atom corresponds to a valuation (*the* valuation in which it is true). For $\theta \in SL$ let

$$S_\theta = \{ \ \alpha \in At^L \mid \alpha \models \theta \ \},$$

so S_θ corresponds to the set of valuations in which θ is true.

For $\vec{s} = s_0, s_1, ..., s_m$ disjoint, nonempty[2], subsets of At^L we define a binary relation $\mid\!\sim_{\vec{s}}$ by

$$\begin{array}{lll} \theta \mid\!\sim_{\vec{s}} \varphi & \iff & \text{either } s_i \cap S_\theta = \emptyset \text{ for all } i = 0, 1, ..., m \\ & & \text{or } s_i \cap S_\theta \neq \emptyset \text{ for some } i = 0, 1, ..., m \\ & & \text{and \ for the least such } i, \ s_i \cap S_\theta \subseteq S_\varphi. \end{array}$$

The idea here is that the atoms in s_0, that is those most highly *ranked*, represent the most normal worlds, the atoms in s_1 the next most normal and so on. Thus $\theta \mid\!\sim_{\vec{s}} \varphi$ amounts to saying that in all the most normal worlds in which θ holds φ also holds. We shall refer to $\mid\!\sim_{\vec{s}}$ as

[2] On occasions it will be useful, with explicit mention, to waive this condition. The same definition of $\mid\!\sim_{\vec{s}}$ still yields a rational consequence relation, which indeed would not be altered by simply removing empty entries from amongst $s_0, s_1, ..., s_m$.

$\vec{s}$-consequence. Notice that in this definition we allow that there may be 'impossible' worlds α not in any s_i. Indeed we allow the possibility that $\vec{s}$ is the the null (empty) sequence, in which case $\mid\!\sim_{\vec{s}} = \mid\!\sim_0$.

Each $\vec{s}$-consequence corresponds to a *ranked preferential model* as defined by Lehmann et al in [8]. The following representation theorem comes from that same paper.

Theorem 1 Every $\vec{s}$-consequence relation is a rational consequence relation and conversely.

In [8] Lehmann et al proffer what they refer to as the *rational closure of a conditional knowledge base* K as the minimal, in the sense of being least information, or making fewest additional assumptions, rational consequence relation satisfying K. Viewed as a relation $\mid\!\sim_{\vec{s}}$ the idea is to give each world the highest ranking it could ever possibly get in any rational consequence which satisfies K, so every world is made to be as normal or usual as possible[3]. Precisely, given a conditional knowledge base K define

$$\begin{aligned} t_i &= \bigcup\{s_i \mid \ \mid\!\sim_{\vec{s}} \text{ satisfies } K\} \\ u_i &= t_i - \bigcup\nolimits_{j<i} t_j \end{aligned}$$

for $i = 0, 1, 2, \ldots$. Let $r+1$ be maximal such that $u_r \neq \emptyset$ (it is easy to see that the u_i are disjoint so such an r must exist, indeed it turns out that all the u_j after u_r will all be null). Then $\mid\!\sim_{\vec{u}}$ is the rational closure of K, where $\vec{u} = u_0, u_1, \ldots, u_r$. The key property of $\mid\!\sim_{\vec{u}}$ is that it does satisfy K, because from this and its construction it follows that as required $\mid\!\sim_{\vec{u}}$ is the rational consequence relation satisfying K in which every world is as normal as it can be in *any* rational consequence relation satisfying K. In what follows it will be assumed that $\vec{u}$ is as above.

As already remarked property inheritance to exceptional subclasses may fail for the rational closure, a specific instance of this being given by the Swedes example. Here the language has propositional variables $\{s, t, f\}$ (standing for 'Swede', 'tall', 'fair' respectively) and

$$K = \{\ s \mid\!\sim t, s \mid\!\sim f\ \},$$

[3] Actually Lehmann et al [8] give a slightly different presentation of the rational closure. The equivalence with the formulation used in this paper is pointed out explicitly in [4].

representing the knowledge that Swedes are normally tall and Swedes are normally fair. For the rational closure, $\mid\!\sim_{\vec{u}}$, of K, $\vec{u} = u_0, u_1$ turns out to be given by

$$u_0 = \{s \wedge t \wedge f,\ \neg s \wedge t \wedge f,\ \neg s \wedge t \wedge \neg f,\ \neg s \wedge \neg t \wedge f, \neg s \wedge \neg t \wedge \neg f\},$$
$$u_1 = \{s \wedge t \wedge \neg f,\ s \wedge \neg t \wedge f,\ s \wedge \neg t \wedge \neg f\}.$$

From this we can see that, as far as $\mid\!\sim_{\vec{u}}$ is concerned, Swedes who are neither tall nor fair ($s \wedge \neg t \wedge \neg f$) are no more exceptional than Swedes who are not tall but are fair ($s \wedge \neg t \wedge f$).

A second approach to characterizing rational consequence relations was given by Pearl in [10] following much earlier work of Adams, [1]. Let $\mathbb{R}^*$ be some proper, countably saturated[4], extension of the usual real closed field, $\mathbb{R}$ and let $[0,1]^* = \{x \in \mathbb{R}^* | 0 \leq x \leq 1\}$.

An *ϵ-probability function* is a function $w : SL \to [0,1]^*$ satisfying, for $\theta, \phi \in SL$,

(P1) $\models \theta \Rightarrow w(\theta) = 1$.
(P2) If $\models \neg(\theta \wedge \phi)$ then $w(\theta \vee \phi) = w(\theta) + w(\phi)$.

As one would anticipate from the name an ϵ-probability function w satisfies all the standard properties associated with a probability function (see for example [9], p13) except of course that the 'probabilities' $w(\theta)$ are in $\mathbb{R}^*$ rather than $\mathbb{R}$. In particular

$$w(\theta) = \sum_{\alpha \in S_\theta} w(\alpha)$$

and w is uniquely determined by the vector of its values on the atoms.

Given an ϵ-probability function w we can define a rational consequence relation $\mid\!\sim_w$ by

$$\theta \mid\!\sim_w \phi \iff \begin{array}{l} \text{either } w(\theta) = 0 \text{ or} \\ w(\theta) \neq 0 \text{ and } w(\neg\phi|\theta) \text{ is 0 or infinitesimal.} \end{array}$$

It is well known that every rational consequence relation (for finite L etc.) is of the form $\mid\!\sim_w$ except the trivial rational consequence relation, $\mid\!\sim_0$. Indeed for this purpose we can restrict the range of w to the set of

[4] This will be needed later. For the reader unfamiliar with this notion it is enough to take $\mathbb{R}^*$ to be simply a proper ultrapower of $\mathbb{R}$.

numbers of the form $p(\epsilon)/q(\epsilon)$ where $\epsilon \in \mathbb{R}^*$ is positive infinitesimal (i.e. $0 < \epsilon < a$ for all $a \in \mathbb{R}, a > 0$) and $p(x), q(x)$ are polynomials over $\mathbb{R}$, $q(x)$ not identically zero.

The idea in Goldzsmidt et al, [5] (though it is formulated slightly differently there) is to consider the set of constraints K_ϵ on an ϵ-probability function w given by

$$K_\epsilon = \{\ w(\theta \wedge \neg\phi) \leq \epsilon w(\theta) \mid \theta \mathrel{|\!\sim} \phi \in K\ \}$$

and take the (unique) solution w to these constraints for which the *entropy*,

$$-\sum_\alpha w(\alpha) \log w(\alpha),$$

is maximal. The resulting rational consequence relation $\mathrel{|\!\sim}_w$ clearly satisfies K (because K_ϵ forces $w(\theta) = 0$ or $w(\neg\phi|\theta) = w(\theta \wedge \neg\phi)/w(\theta) \leq \epsilon$) and is what we have been referring to above as the *maximum entropy closure* of K.

The approach as we have described it here however fudges a couple of difficulties which seem to have gone largely unnoticed (Weydert, [14], is an exception). Firstly, we do not have the function log in $\mathbb{R}^*$ so what does it mean to *maximize the entropy*? Well, of course, we could just add log, and let us say also *exp*, to $\mathbb{R}$ in the first place and then carry them over to $\mathbb{R}^*$.

A second problem now arises however, indeed it has been there all along but was just not relevant before. The problem is, how do we know that this maximum entropy closure does not depend on our initial choice of positive infinitesimal ϵ? In the simple situation such as that described in [10] we could, as already indicated above, have used the field of rational polynomials in ϵ over $\mathbb{R}$ in place of $\mathbb{R}^*$ and it is quite easy to see that such fields are the same up to isomorphism for any choice of positive infinitesimal ϵ. But that is far from clear when we also allow in the functions log and *exp*. Fortunately, by a deep result of Wilkie, [15], the exponential field of the reals is o-minimal and it is a corollary of this that the particular choice of ϵ is irrelevant. We shall shortly state a generalization of this result.

Turning now to the particular maximum entropy solution, w_G say, given by the Goldzsmidt et al approach, in the Swedes case it turns out that $w_G(s \wedge \neg t \wedge f)$ and $w_G(s \wedge t \wedge \neg f)$ both have order ϵ whereas

$w_G(s \wedge \neg t \wedge \neg f)$ has order ϵ^2, giving the 'desired' conclusion that

$$s \wedge \neg t \mid\!\sim_{w_G} f.$$

Our 'objection' to the standard interpretation of this result, that it sanctions (at least in this instance) this $\mid\!\sim_{w_G}$ as the simplest, least prejudiced, rational consequence relation satisfying K and thence in turn 'property inheritance to exceptional subclasses', goes right back to the definition of $\mid\!\sim_w$. Putting aside the case when $w(\theta) = 0$, corresponding to θ being judged impossible, the intention (we believe) is that ϕ should normally follow from θ if the probability of $\neg\phi$ given θ is *infinitesimal in relation to the probability of* θ. This has conventionally been interpreted, as in the definition, as

> $w(\neg\phi|\theta)$ *is infinitesimal when viewed from the perspective of the standard universe* $\mathbb{R}$.

However we would argue that the 'correct' interpretation is

> $w(\neg\phi|\theta)$ *is infinitesimal when viewed from the perspective of the smallest universe containing* $w(\theta)$.

With this interpretation $w_G(s \wedge \neg t \wedge \neg f)$ is no longer infinitesimal in comparison with $w_G(s \wedge \neg t)$ since ϵ^2 clearly lies in any subfield of $\mathbb{R}^*$ containing ϵ. To put it more picturesquely, if you lived in a world in which a probability of size ϵ was 'real' then a probability of size ϵ^2 would be equally 'real', and certainly not infinitesimal from your point of view. In the next section we develop a formulation of rational consequence along these lines. In turn this will lead to a new version of the 'maximum entropy closure' of a conditional knowledge base K.

A New Maximum Entropy Approach

In this section we will define our revised ME-approach, which uses the full power of infinitesimals and defines a much stronger ordering relation from which to draw inferences. Roughly, given positive infinitesimals ϵ and δ we will take ϵ to be 'smaller' than δ if, no matter which real function we take, if we apply that function to δ then the result will be larger than ϵ, or ≤ 0. So we can never use a real function on δ to produce a positive infinitesimal smaller than ϵ. In particular, our system would reckon ϵ^2 to be of comparable 'size' to ϵ.

Specifically, we define $\mathcal{F}_n$ to be the set of all n-ary functions which are first-order definable in $\mathbb{R}_{exp}$, the field of reals augmented with the functions *exp* and *log*, extending them to a nonstandard, countably saturated, extension $\mathbb{R}^*_{exp}$ of $\mathbb{R}_{exp}$ in the obvious way via their defining formulae. Notice that we may assume that the aforementioned $\mathbb{R}^*$ is the reduct of $\mathbb{R}^*_{exp}$ to the usual language of fields so that in particular they share the same universe.

For $0 < \gamma, \delta \in \mathbb{R}^*$ we say that γ is *infinitely smaller* than δ, $\gamma \ll \delta$ (or $\delta \gg \gamma$), if for all $f \in \mathcal{F}_1$, $\gamma < f(\delta)$ or $f(\delta) \leq 0$. We say that γ is *comparable* to δ, $\gamma \sim \delta$, if $\gamma \not\ll \delta$ and $\delta \not\ll \gamma$.

Notice then that $0 < \epsilon \in \mathbb{R}^*$ is infinitesimal just if $\epsilon \ll 1$ and if $0 < \delta, \gamma$ then $\delta + \gamma \sim \max\{\delta, \gamma\}$ and $\delta\gamma \sim \min\{\delta, \gamma\}$. Notice also that because of the countable saturation if $0 < \gamma \ll \delta \leq 1$ then there are infinitesimals ν, τ such that $0 < \tau \ll \gamma \ll \nu \ll \delta$.

Theorem 2 The relation $\sim$ is an equivalence relation on the (positive) infinitesimals in $\mathbb{R}^*$. Furthermore if $0 < \delta_1, \delta_2, \nu_1, \nu_2$ are infinitesimal and $\delta_1 \sim \delta_2$ and $\nu_1 \sim \nu_2$ then

$$\delta_1 \ll \nu_1 \iff \delta_2 \ll \nu_2.$$

The proof of this result, together with all the other new results in this paper, can be found in the appendix.

We now fulfill the intentions outlined above by defining, for $w : SL \to \mathbb{R}^*$ an ϵ-probability function, a binary consequence relation on $SL \times SL$ by

$$\begin{aligned} \theta \mid\!\mid\!\!\sim_w \phi \quad \iff \quad & \text{either } w(\theta) = 0 \text{ or} \\ & w(\theta) \neq 0 \text{ and } w(\theta \wedge \neg\phi) \ll w(\theta). \end{aligned}$$

Theorem 3 For $w : SL \to \mathbb{R}^*$ an ϵ-probability function, $\mid\!\mid\!\!\sim_w$ is a non-trivial rational consequence relation. Conversely every non-trivial rational consequence relation is of the form $\mid\!\mid\!\!\sim_w$ for some such ϵ-probability function w.

Continuing along the path outlined earlier it seems we should now be interested in the ϵ-probability function w of maximum entropy for which

$|\!\!\sim_w$ satisfies K. Unfortunately, except in rather trivial cases, such an object does not exist. The problem is that given w for which $|\!\!\sim_w$ satisfies K we can always find a w' of higher entropy with $|\!\!\sim_{w'}$ also satisfying K. To see this let $1 = \gamma_0 \gg \gamma_1 \gg \gamma_2 \gg ... \gg \gamma_m$ be such that for each $\alpha \in At^L$ either $w(\alpha) = 0$ or there is some $0 \leq j \leq m$ with $w(\alpha) \sim \gamma_j$. Let $1 = \gamma_0 \gg \delta_1 \gg \gamma_1 \gg \delta_2 \gg \gamma_2 \gg ... \gg \delta_m \gg \gamma_m$, define $u : At^L \to \mathbb{R}^*$ by

$$u(\alpha) = \begin{cases} w(\alpha) & \text{if } w(\alpha) \sim 1 \text{ or } w(\alpha) = 0, \\ \delta_i & \text{if } w(\alpha) \sim \gamma_i,\ i > 0, \end{cases}$$

and define the ϵ-probability function w' to be the *normalization* of u, i.e.

$$w'(\alpha) = \frac{u(\alpha)}{\sum_{\beta \in At^L} u(\beta)} \ .$$

Then $|\!\!\sim_{w'}$ still satisfies K and, provided not all the non-zero $w(\alpha) \sim 1$, will have strictly higher entropy.

Clearly from this example what we should do is to somehow quotient, or filter, out the differences between the $1, \gamma_1, \gamma_2, ..., \gamma_m$ and the $1, \delta_1, \delta_2, ..., \delta_m$. What should matter is the relative (with respect to $\ll$ and $\sim$) sizes of the infinitesimals within these sequences.

This is of course also a difficulty with the Goldzsmidt et al approach and we shall address it in a similar way by considering the maximum entropy solution of a set of constraints analogous to K_ϵ rather than attempting to take the maximum entropy solution over *all* ϵ-probability functions w for which $|\!\!\sim_w$ satisfies K.

Let B be a sufficiently large set of (positive) infinitesimals, by which we mean that B contains at least $|K|$ incomparable elements (with respect to $\sim$). A *weighting* (of K to B) is a function $\sigma : K \to B$. For such a σ let

$$K_\sigma = \{\ w(\theta \wedge \neg\phi) \leq \sigma(\theta \mid\!\sim \phi) w(\theta) \mid \theta \mid\!\sim \phi \in K\ \}$$

and let M_σ be the maximum entropy solution to K_σ. [Because this can all be expressed in $\mathbb{R}^*_{exp}$ this solution is well defined.] It is easy to see that since M_σ satisfies K_σ the rational consequence relation $\mid\!\sim_{M_\sigma}$ satisfies K. However, because of the enhanced notion of 'infinitesimal with respect to' demanded by $|\!\!\sim_{M_\sigma}$ this rational consequence relation may not satisfy K. If it does we say that the weighting σ is *valid*.

Our next theorem shows that by limiting ourselves to weightings on such a set B we effect the required 'filtering'.

Theorem 4 Suppose B, C are sufficiently large for K and σ, τ are weightings of K to B, C respectively such that for $\theta \mid\!\sim \phi, \psi \mid\!\sim \eta$ in K

$$\sigma(\theta \mid\!\sim \phi) \lesseqgtr \sigma(\psi \mid\!\sim \eta) \iff \tau(\theta \mid\!\sim \phi) \lesseqgtr \tau(\psi \mid\!\sim \eta).$$

Then $|\mid\!\sim_{M_\sigma} = |\mid\!\sim_{M_\tau}$.

With this in place we can proceed with our plan and take as the correct(!) maximum entropy closure of K that one of the (finitely) many $|\mid\!\sim_{M_\sigma}$, with σ a valid weight of K to B, whose entropy is highest. Our expectation that this simplest, minimum information, solution to K should equal the rational closure of K is fulfilled by the following theorem.

Theorem 5 Let B be sufficiently large for K and σ a valid weighting of K to B such that amongst all the valid weightings of K to B M_σ has the greatest entropy. Then $|\mid\!\sim_{M_\sigma}$ is the rational closure of K.

In particular then, in the case of the Swedes knowledge base $K = \{ s \mid\!\sim t,\ s \mid\!\sim f \}$ the M_σ which maximizes the entropy for any such weighting does not sanction $s \wedge \neg t \mid\!\sim f$, that even short Swedes are normally fair. For those who would see this as a sad loss there is a modicum of comfort in the following observation. If instead of this knowledge base K we had taken the equivalent knowledge base $K' = \{ s \mid\!\sim t \wedge f \}$, *equivalent* in the sense that it is satisfied by exactly the same rational consequence relations as K (and hence has the same rational closure) then the Goldzsmidt et al maximum entropy solution no longer satisfies $s \wedge \neg t \mid\!\sim f$.

We end this section by giving a relatively simple characterization of those rational consequence relations satisfying K which are of the form $|\mid\!\sim_{M_\sigma}$ for some weighting σ of K to a sufficiently large B.

Given $\alpha \in At^L$ say that α *refutes* the conditional $\theta \mid\!\sim \phi$ if $\alpha \in S_{\theta\wedge\neg\phi}$. Now let $\vec{s} = s_0, s_1, s_2, ..., s_m$ be, as usual, non-empty disjoint subsets of At^L. We shall say that $\vec{s}$ provides a *vindicated* solution of K if $\mid\!\sim_{\vec{s}}$ satisfies K and the following two conditions hold:

(a) $\bigcup_i s_i = \bigcup_i u_i$, where, as above, $\vec{u}$ gives the rational closure of K.
(b) Whenever $\alpha \in \bigcup_i s_i$ then either α refutes no conditional in K and $\alpha \in s_0$ or α refutes some conditional $\theta \mid\!\sim \phi$ in K and no $\alpha' \in \bigcup_{j<i} s_j$ refutes that conditional.

Theorem 6 If $\vec{s}$ is a vindicated solution of K and B is sufficiently large for K then there is valid weighting of K to B such that $|\!\!\sim_{M_\sigma} = \,\mid\!\sim_{\vec{s}}$. Conversely if σ is a valid weighting of K to B then there is a vindicated solution $\vec{s}$ of K such that $|\!\!\sim_{M_\sigma} = \,\mid\!\sim_{\vec{s}}$.

We shall refer to the binary relation on $SL \times SL$ which holds between θ and ϕ just if $\theta \mid\!\sim_{\vec{s}} \phi$ for all vindicated solutions $\vec{s}$ of K as the *vindicated closure* of K.

It is easy to see that in general the vindicated closure of K will not be a rational consequence relation. In the next section we point out an alternative (and earlier) characterization of this same relation due to Bamber et al.

Bamber et al's Entailment with Near Surety

In [2], Bamber and Goodman, and in [3], Bamber, describe, amongst much other related work, the notion of *strong entailment with near surety.* Setting Bamber et al's work within the context of this paper let $0 < \nu \ll \xi \ll 1$, let $\rho : K \to (0, \nu]^*$ and let X_ρ denote the set of ϵ-probability functions satisfying K_ρ. Since K is consistent and the constraints in K_ρ are linear this set is non-empty and convex. Thus it makes sense to talk about *the uniform normalized measure,* Υ_ρ, *on* X_ρ since this measure is definable (with parameters from the range of ρ) in $\mathbb{R}^*$.

Rephrased in the style of this paper Bamber et al define a binary relation $\mid\!\approx_\rho$ on $SL \times SL$ by

$$\theta \mid\!\approx_\rho \phi \iff \Upsilon_\rho\{w \in X_\rho | w(\theta \wedge \neg\phi) \leq \xi w(\theta)\} \geq 1 - \xi.$$

Loosely (indeed precisely as it turns out), $\theta \mid\!\approx_\rho \phi$ holds just if for all except possibly an infinitesimal fraction of the $w \in X_\rho$, $\theta \mid\!\sim_w \phi$. The next theorem will clarify this rather cryptic parenthetical remark, but first a definition.

Let K, ν, ξ be as above. Bamber et al say that K *strongly entails* $\theta \mid\!\sim \phi$ *with near surety* if $\theta \mid\!\approx_\rho \phi$ holds for all $\rho : K \to (0, \nu]^*$.

Theorem 7 For K, ν, ξ etc as above $\mid\!\approx_\rho$ is independent of the choice of ξ. Furthermore, whether or not K strongly entails $\theta \mid\!\sim \phi$ with near surety is independent of the choice of ν, ξ (subject to $0 < \nu \ll \xi \ll 1$).

The connection of this notion to the results in this paper is given by the following theorem.

Theorem 8 A finite consistent conditional knowledge base K strongly entails $\theta \mid\!\sim \phi$ with near surety just if $\theta \mid\mid\!\sim_{M_\sigma} \phi$ for all valid weightings of K to B, where B is sufficiently large for K. Equivalently, just if $\theta \mid\!\sim_{\vec{s}} \phi$ holds for all vindicated solutions $\vec{s}$ of K.

In other words the vindicated closure of K is the same as the closure of K based on strong entailment with near surety.

This is a somewhat surprising result for the following reason. Suppose we view the vindicated closure of K as the intersection of all the relations $\mid\mid\!\sim_{M_\sigma}$ which satisfy K, for some B sufficiently large for K. Then clearly we could replace B by $(0, \nu]^*$ for ν as above without altering the resulting closure. [We cannot also replace $\mid\mid\!\sim_{M_\sigma}$ by $\mid\!\sim_{M_\sigma}$ however, as can be easily checked by considering the Swedes knowledge base $K = \{s \mid\!\sim t,\ s \mid\!\sim f\ \}$.] In a sense then Theorem 8 says that we get the same answers if we pick as representative of the X_σ either the centers of mass according to the uniform measure, and use the $\mid\!\sim_w$, or the points of maximum entropy, and use the $\mid\mid\!\sim_w$ satisfying K. Furthermore, as Bamber et al observe in 4.2 of [2] the use of a uniform measure is not essential, substitution of any of a wide range of non-uniform measures would not change the results, whilst it is clear from our proofs that they will equally go through with a range of 'information measures' in place of Shannon's. It seems then that what we have called here the vindicated closure has some claim to an 'established presence'.

Conclusion

We have developed a new, more intuitive approach to maximum entropy inference which uses the full weight of the infinitesimal calculus, and thus avoids the need to choose an arbitrary infinitesimal in order to make calculations. Inferences are drawn independently of the particular infinitesimals used, and depend only on the relative sizes of those we choose. Moreover, our system shows that the maximum entropy inference process and the rational closure of Lehmann et al yield the same rational consequence relation. As ME is widely viewed as the least biased inference process, this provides a further justification for taking the rational closure as the fairest, least assumptive rational consequence relation satisfying a knowledge base. In addition, the family of rational consequence relations we describe provides an alternative derivation of Bamber et al's notion of strong entailment with near surety, thus further justifying the use of our revised approach.

Appendix

In this appendix we give the proofs of the (new) results stated in the body of the paper. Before that however the following lemma will prove invaluable.

Lemma 9 Suppose that $\tau_1, \tau_2, ..., \tau_r \in \mathbb{R}^*_{exp}$ and $f \in \mathcal{F}_{r+1}$. Then there are some $f_1, f_2, ..., f_k \in \mathcal{F}_r$ such that

$$-\infty < f_1(\tau_1, ..., \tau_r) < f_2(\tau_1, ..., \tau_r) < ... < f_k(\tau_1, ..., \tau_r) < \infty$$

and on each of the intervals $(-\infty, f_1(\tau_1, ..., \tau_r)), (f_k(\tau_1, ..., \tau_r), \infty)$, $(f_i(\tau_1, ..., \tau_r), f_{i+1}(\tau_1, ..., \tau_r)), i = 1, ..., k-1$, the function $f(x, \tau_1, \tau_2, ..., \tau_r)$ is continuous, either constant or strictly monotone, and either non-negative or non-positive.

Proof. This lemma follows from results of Wilkie, [15] and van den Dries, [13], in particular the *Cell Decomposition Theorem* from that second publication. For the fine details we refer the reader to [6]. ■

Proof of Theorem 2. That $\sim$ is an equivalence relation follows from the second part. To show the second part suppose that $\delta_1 \sim \delta_2$ and $\nu_1 \sim \nu_2$ and $\delta_1 \ll \nu_1$ but $\delta_2 \not\ll \nu_2$. Then by Lemma 9 there are $f_1, f_2, f_3 \in \mathcal{F}_1$ such that $f_1(\delta_2) \leq \delta_1$, $f_2(\nu_1) \leq \nu_2$ and $f_3(\nu_2) \leq \delta_2$. Further, since the other possibilities can be ruled out or replaced, there must be $r_i > 0$, $r_i \in \mathbb{R}$ for $i = 1, 2, 3$ such that each f_i is positive and decreasing on $(0, r_i)$. From this it follows that $0 < f_1 f_3 f_2(\nu_1) \leq \delta_1$, contradicting $\delta_1 \ll \nu_1$. The result follows. ■

Proof of Theorem 3. In view of Theorem 1 it is sufficient to prove that for every ϵ-probability function $w : SL \to \mathbb{R}^*$, $|\!\sim_w = |\!\sim_{\vec{s}}$ for some non-null $\vec{s}$, and conversely.

Suppose first that w is an ϵ-probability function and let $\alpha_0, \alpha_1, ..., \alpha_m \in At^L$ be such that

$$1 \sim w(\alpha_0) \gg w(\alpha_1) \gg ... \gg w(\alpha_m) > 0$$

and for all $\alpha \in At^L$, either $w(\alpha) = 0$ or $w(\alpha) \sim w(\alpha_i)$ for some $i = 0, 1, ..., m$. Notice that since the sum of the $w(\alpha)$ is 1 some such α_0 must exist. Now define $\vec{s} = s_0, s_1, \ldots, s_m$ by

$$s_i = \{\alpha \in At^L \mid w(\alpha) \sim w(\alpha_i)\}.$$

Then for $\theta, \phi \in SL$,

$$w(\theta) = 0 \iff s_i \cap S_\theta = \emptyset \text{ for all } i = 0, 1, ..., m \tag{1}$$

and

$$w(\theta \wedge \neg\phi) \ll w(\theta) \iff \text{for some } i = 0, 1, ..., m, \quad S_\theta \cap \bigcup_{j \le i} s_j \neq \emptyset,\ S_{\theta \wedge \neg\phi} \cap \bigcup_{j \le i} s_j = \emptyset, \tag{2}$$

from which the desired conclusion follows.

In the other direction let $\mid\!\sim_{\vec{s}}$ be non-trivial, so $\vec{s} = s_0, s_1, ..., s_m$, let

$$1 = \gamma_0 \gg \gamma_1 \gg ... \gg \gamma_m > 0$$

and define the ϵ-probability function w to be the normalization of the function $u : At^L \to [0, 1]^*$ defined by

$$u(\alpha) = \begin{cases} \gamma_i & \text{if } \alpha \in s_i, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that since $\vec{s}$ is not null the sum of the $u(\alpha)$ is greater than 0 and w is well defined. With this definition (1) and (2) again hold and consequently $|\mid\!\sim_w = \mid\!\sim_{\vec{s}}$, as required. ■

Before proving Theorem 4 we will need several lemmas which are of some interest in their own right.

Lemma 10 Let $0 < \gamma_1 \ll \gamma_2 \ll ... \ll \gamma_r \ll 1$ and $f \in \mathcal{F}_r$. Then either $f(\gamma_1, \gamma_2, ..., \gamma_r)$ is not positive infinitesimal or $f(\gamma_1, \gamma_2, ..., \gamma_r) \sim \gamma_i$ for some $i = 1, 2, ..., r$.

Proof. The proof is by induction on r. Assume the result holds below r but that $f(\gamma_1, \gamma_2, ..., \gamma_r)$ is positive infinitesimal and not comparable to any of $\gamma_1, \gamma_2, ..., \gamma_r$. By Lemma 9 (or directly from the definition of o-minimality!) the set of x_1 such that

$$f(\gamma_1, \gamma_2, ..., \gamma_r) = f(x_1, \gamma_2, ..., \gamma_r)$$

is the union of a finite set of points and open intervals on which $f(x_1, \gamma_2, ..., \gamma_r)$ is constant. In turn applying Lemma 9 to the function $f^*(x_1, \gamma_2, ..., \gamma_r)$ which takes value 1 if x_1 is in an open interval of constancy of $f(x_1, \gamma_2, ..., \gamma_r)$ and zero otherwise we see that there can only be finitely many such intervals and their midpoints must be definable from $\gamma_2, \gamma_3, ..., \gamma_r$. By the inductive hypothesis then it is impossible for γ_1 to lie in one of these intervals, otherwise $f(\gamma_1, \gamma_2, ..., \gamma_r)$ would be definable from $\gamma_2, \gamma_3, ..., \gamma_r$ and the midpoint of that interval, and hence, by the inductive hypothesis, would be comparable to one of $\gamma_2, \gamma_3, ..., \gamma_r$. It follows that γ_1 must be definable from $\gamma_2, \gamma_3, ..., \gamma_r, f(\gamma_1, \gamma_2, ..., \gamma_r)$. If $r = 1$ this is where the proof ends, since we have $\gamma_1 \sim f(\gamma_1)$. Otherwise we have

$$\gamma_1 = g(\gamma_2, \gamma_3, ..., \gamma_r, f(\gamma_1, \gamma_2, ..., \gamma_r))$$

for some $g \in \mathcal{F}_r$.

From this it follows (by transposing γ_1 and $f(\gamma_1, \gamma_2, ..., \gamma_r)$ if necessary) that we may have assumed at the start that

$$0 < f(\gamma_1, \gamma_2, ..., \gamma_r) \ll \gamma_1.$$

By Lemma 9 there is a subinterval of $(0, 1)$,

$$(h_1(\gamma_1, ..., \gamma_{r-1}), h_2(\gamma_1, ..., \gamma_{r-1})),$$

with $h_1, h_2 \in \mathcal{F}_{r-1}$, containing γ_r, on which $f(\gamma_1, \gamma_2, ..., \gamma_{r-1}, x_r)$ is positive, strictly monotone or constant. By the inductive hypothesis we have that

$$h_1(\gamma_1, ..., \gamma_{r-1}), h_2(\gamma_1, ..., \gamma_{r-1}) \not\sim \gamma_r$$

so that

$$h_1(\gamma_1, ..., \gamma_{r-1}) + \gamma_1 < \gamma_r < h_2(\gamma_1, ..., \gamma_{r-1}) - \gamma_1.$$

From the monotonicity of $f(\gamma_1, \gamma_2, ..., \gamma_{r-1}, x_r)$ on this interval then at least one of

$$f(\gamma_1, \gamma_2, ..., \gamma_{r-1}, h_1(\gamma_1, ..., \gamma_{r-1}) + \gamma_1),$$

$$f(\gamma_1, \gamma_2, ..., \gamma_{r-1}, h_2(\gamma_1, ..., \gamma_{r-1}) - \gamma_1),$$

must be positive, infinitesimal and less than $f(\gamma_1, \gamma_2, ..., \gamma_{r-1}, \gamma_r)$. But this is impossible by the inductive hypothesis. ■

Lemma 11 Let σ be a weighting for K, not necessarily valid, and let w be an ϵ-probability function which satisfies K_σ. Then for $\alpha \in At^L$, $w(\alpha) \lessapprox M_\sigma(\alpha)$

Proof. Enumerate the atoms of At^L as $\alpha_1, \alpha_2, ..., \alpha_{2^n}$ and to simplify the notation let

$$\begin{aligned} x_i &= M_\sigma(\alpha_i), \\ y_i &= w(\alpha_i). \end{aligned}$$

Suppose now on the contrary to the lemma that $y_t \gg x_t$ for some t. Let δ be such that $x_t \ll \delta \ll y_t$ and let

$$\vec{z} = (1-\delta)\vec{x} + \delta\vec{y}.$$

Clearly the ϵ-probability function corresponding to $\vec{z}$ satisfies K_σ so we shall have our contradiction once we have shown that

$$-\sum_{i=1}^{2^n} z_i \log z_i > -\sum_{i=1}^{2^n} x_i \log x_i.$$

First suppose that $i \neq t$ and

$$-z_i \log z_i + x_i \log x_i < 0.$$

If x_i is infinitesimal then z_i is also infinitesimal because δ is. Since $-x \log x$ is increasing on the infinitesimals, it must be the case that $z_i < x_i$, so $y_i < x_i$. Hence since log is increasing,

$$\begin{aligned} -z_i \log z_i + x_i \log x_i &\geq (-z_i + x_i) \log x_i \\ &= \delta(x_i - y_i) \log x_i \\ &\geq \delta x_i \log x_i \\ &> -\delta. \end{aligned}$$

So in this case the loss of entropy derived from this term resulting from the shift from $\vec{x}$ to $\vec{z}$ is less than δ. Now suppose that x_i is not infinitesimal. Then neither is z_i. We cannot use the same argument as above, because $-x \log x$ is decreasing on (e^{-1}, ∞). However we can appeal to the mean value theorem to show that there exists c_i lying between x_i and z_i for which

$$\frac{-z_i \log z_i + x_i \log x_i}{z_i - x_i} = -c_i - \log c_i$$

and so

$$-z_i \log z_i + x_i \log x_i = \delta(y_i - x_i)(-c_i - \log c_i).$$

Now c_i is non-infinitesimal so $-\log c_i$ is finite, and hence $(y_i - x_i)(-c_i - \log c_i)$ must be finite. So in this case similarly any loss in entropy must be at most a finite multiple of δ.

Thus our total loss of entropy for atoms other than that corresponding to α_t must be less than $a\delta$ for some $a \in \mathbb{R},\ a > 0$. Now consider

$$z_t \log z_t - x_t \log x_t.$$

Since $x_t \ll \delta$ we know that $-x_t \log x_t \ll \delta$. So the negative part of this expression is certainly less than δ. Now

$$-z_t \log z_t = -((1-\delta)x_t + \delta y_t)\log((1-\delta)x_t + \delta y_t) > -\delta y_t \log \delta y_t,$$

because $z_t, \delta y_t$ are infinitesimal, and by Lemma 10

$$(-\log \delta y_t)^{-1} \sim \delta y_t \ll y_t.$$

Hence

$$-\delta y_t \log \delta y_t > (a+1)\delta.$$

From this it follows that $-z_t \log z_t > (a+1)\delta$ and hence the net difference in entropy between $\vec{z}$ and $\vec{x}$ is positive, which gives the required contradiction. ■

Lemma 12 Let $0 < \gamma_1 \ll \gamma_2 \ll ... \ll \gamma_r \ll 1$ and $0 < \delta_1 \ll \delta_2 \ll ... \ll \delta_r \ll 1$, and let $\Psi(x_1, x_2, ..., x_r)$ be a formula of the language of the structure $\mathbb{R}^*_{exp}$. Then

$$\mathbb{R}^*_{exp} \models \Psi(\gamma_1, \gamma_2, ..., \gamma_r) \iff \mathbb{R}^*_{exp} \models \Psi(\delta_1, \delta_2, ..., \delta_r).$$

Proof. By applying the result twice if necessary we may assume that $\delta_r \ll \gamma_1$. Let $f(x_1, x_2, ..., x_r)$ be the characteristic function of $\Psi(x_1, x_2, ..., x_r)$ and let $1 \leq j \leq r$. By Lemma 9 there are some $f_1, f_2, ..., f_k \in$

$\mathcal{F}_{r-1}$ such that $f(\delta_1, \delta_2, ..., \delta_{j-1}, x_j, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r)$ is constant on each of the intervals

$$(-\infty, f_1(\delta_1, \delta_2, ..., \delta_{j-1}, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r)),$$
$$(f_i(\delta_1, \delta_2, ..., \delta_{j-1}, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r), f_{i+1}(\delta_1, \delta_2, ..., \delta_{j-1}, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r)),$$
$$(f_r(\delta_1, \delta_2, ..., \delta_{j-1}, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r), \infty)$$

where $i = 1, 2, ..., k$. By Lemma 10, for one of these intervals, (a_1, a_2) say, we must have $a_1 \lessapprox \delta_{j-1} < \delta_j, \gamma_j < \gamma_{j+1} \lessapprox a_2$. Hence

$$f(\delta_1, \delta_2, ..., \delta_{j-1}, \delta_j, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r) = \\ f(\delta_1, \delta_2, ..., \delta_{j-1}, \gamma_j, \gamma_{j+1}, \gamma_{j+2}, ..., \gamma_r).$$

Given these identities for $j = 1, 2, ..., r$ it follows that

$$f(\delta_1, \delta_2, ..., \delta_r) = f(\gamma_1, \gamma_2, ..., \gamma_r),$$

as required. ■

Proof of Theorem 4. First suppose that the values $\sigma(\theta \mid\!\sim \phi)$, $\theta \mid\!\sim \phi \in K$ are pairwise incomparable and similarly for the $\tau(\theta \mid\!\sim \phi)$, $\theta \mid\!\sim \phi \in K$. Then since M_σ is definable in $\mathbb{R}^*_{exp}$ from K_σ etc., by Lemma 12, for $\alpha, \beta \in At^L$, $\theta \mid\!\sim \phi \in K$,

$$\begin{aligned} M_\sigma(\alpha) = 0 &\iff M_\tau(\alpha) = 0, \\ M_\sigma(\alpha) \ll M_\sigma(\beta) &\iff M_\tau(\alpha) \ll M_\tau(\beta), \\ M_\sigma(\alpha) \sim \sigma(\theta \mid\!\sim \phi) &\iff M_\tau(\alpha) \sim \tau(\theta \mid\!\sim \phi), \end{aligned} \tag{3}$$

[Actually we cannot apply Lemma 12 *directly* here since the relation $\ll$ is not first order definable in $\mathbb{R}^*_{exp}$. However if, say, $M_\sigma(\alpha) \not\ll M_\sigma(\beta)$ then there is an $f \in \mathcal{F}_1$ such that $0 < f(M_\sigma(\beta)) \leq M_\sigma(\alpha)$ and Lemma 12 *does apply* to this expression.]

In the case where, say, not all the values $\sigma(\theta \mid\!\sim \phi)$, $\theta \mid\!\sim \phi \in K$ are pairwise incomparable define weightings $\sigma', \sigma'' : K \to B$ by

$$\begin{aligned} \sigma'(\theta \mid\!\sim \phi) &= \min\{\gamma \in B \mid \gamma \sim \sigma(\theta \mid\!\sim \phi)\}, \\ \sigma''(\theta \mid\!\sim \phi) &= \max\{\gamma \in B \mid \gamma \sim \sigma(\theta \mid\!\sim \phi)\}. \end{aligned}$$

Then by the special case above $|\!\sim_{M_{\sigma'}} = |\!\sim_{M_{\sigma''}}$ and by (3) and Lemma 10 $M_{\sigma'}(\alpha) \sim M_{\sigma''}(\alpha)$ whenever one (hence both) of these is non-zero. Now clearly $M_{\sigma'}$ satisfies K_σ (since $\sigma'(\theta \mid\!\sim \phi) \leq \sigma(\theta \mid\!\sim \phi)$ for $\theta \mid\!\sim \phi \in K$) and M_σ satisfies $K_{\sigma''}$ so by Lemma 11, whenever $M_{\sigma'}(\alpha) > 0$,

$$M_{\sigma'}(\alpha) \lesssim M_\sigma(\alpha) \lesssim M_{\sigma''}(\alpha).$$

Putting these facts together we see that for each α either

$$M_{\sigma'}(\alpha) = M_\sigma(\alpha) = M_{\sigma''}(\alpha) = 0$$

or

$$M_{\sigma'}(\alpha) \sim M_\sigma(\alpha) \sim M_{\sigma''}(\alpha).$$

Hence $|\!\sim_{M_{\sigma'}} = |\!\sim_{M_\sigma} = |\!\sim_{M_{\sigma''}}$ and the theorem now follows. ■

Before proving Theorem 5 we need two lemmas.

Lemma 13 Let σ be a valid weighting for K and $|\!\sim_{\vec{u}}$ the rational closure of K. Then for all $\alpha \in At^L$, exactly one of the following conditions holds.

1. $\alpha \in u_0$ and $M_\sigma(\alpha) \sim 1$.
2. $\alpha \notin \bigcup_i u_i$ and $M_\sigma(\alpha) = 0$.
3. $\alpha \in \bigcup_{i>0} u_i$ and $M_\sigma(\alpha) \sim \min\{\sigma(\theta \mid\!\sim \phi) \mid \alpha \text{ refutes } \theta \mid\!\sim \phi \in K\}$.

Proof. If $\alpha \in u_0$ then the ϵ-probability function w which gives value 1 to α and 0 on all the remaining atoms satisfies K_σ, which, by Lemma 11, forces $M_\sigma(\alpha) \sim 1$.

If $\alpha \notin \bigcup_i u_i$ then, since certainly $|\!\sim_{M_\sigma}$ satisfies K, we must have $M_\sigma(\alpha) = 0$ since otherwise the $\vec{s}$ such that $|\!\sim_{\vec{s}} = |\!\sim_{M_\sigma}$ would have appeared in the 'construction' of $\vec{u}$ and given $\alpha \in \bigcup_i s_i \subseteq \bigcup_i u_i$, contradiction.

To show the last case suppose on the contrary that for some atom β,

$$0 < M_\sigma(\beta) \ll \min\{\sigma(\theta \mid\!\sim \phi) \mid \beta \text{ refutes } \theta \mid\!\sim \phi \in K\}.$$

Now pick $\lambda \in \mathbb{R}^*$ such that

$$M_\sigma(\beta) \ll \lambda \ll \min\{\sigma(\theta \mid\!\sim \phi) | \beta \text{ refutes } \theta \mid\!\sim \phi \in K\}$$

and

$$\lambda \ll \min\{M_\sigma(\alpha) | M_\sigma(\beta) \ll M_\sigma(\alpha)\},$$

and define w to be the ϵ-probability function corresponding to the normalization of the function v defined on the atoms by

$$v(\alpha) = \begin{cases} M_\sigma(\alpha)^2 & \text{if } \alpha \neq \beta, \\ \lambda & \text{if } \alpha = \beta. \end{cases}$$

We shall show that v, and hence w, satisfies K_σ by showing that for $\theta \mid\!\sim \phi \in K$,

$$v(\theta \wedge \neg\phi) \leq \sigma(\theta \mid\!\sim \phi) v(\theta),$$

where $v(\theta) = \sum_{\alpha \in S_\theta} v(\alpha)$ etc..

There are 2 cases. If $\beta \notin S_{\theta\wedge\neg\phi}$ then, since σ is valid (i.e. $\mid\!\sim_{M_\sigma}$ satisfies K_σ), there is $\alpha \in S_\theta$ such that $M_\sigma(\alpha) \gg M_\sigma(\alpha')$ for all $\alpha' \in S_{\theta\wedge\neg\phi}$. Hence for such α', $M_\sigma(\alpha)^3, \sigma(\theta \mid\!\sim \phi) \geq M_\sigma(\alpha')$ and

$$v(\alpha') = M_\sigma(\alpha')^2 \leq \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha') \leq \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha)^3.$$

Summing now gives

$$\begin{aligned} v(\theta \wedge \neg\phi) &\leq |S_{\theta\wedge\neg\phi}| \, \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha)^3 \\ &\leq \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha)^2 \\ &\leq \sigma(\theta \mid\!\sim \phi) v(\theta), \end{aligned}$$

as required.

In the second case, of $\beta \in S_{\theta\wedge\neg\phi}$, we know that $M_\sigma(\beta) \ll \sigma(\theta \mid\!\sim \phi)$ so $v(\beta) = \lambda \ll \sigma(\theta \mid\!\sim \phi)$. Again let α be such that $M_\sigma(\alpha) \gg M_\sigma(\alpha')$ for all $\alpha' \in S_{\theta\wedge\neg\phi}$. Then for such an $\alpha' \neq \beta$,

$$v(\alpha') \leq \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha)^3 \leq \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha) v(\alpha)$$

as before, whilst for $\alpha' = \beta$

$$v(\alpha') = \lambda \ll \sigma(\theta \mid\!\sim \phi), M_\sigma(\alpha),$$

since $M_\sigma(\beta) \ll M_\sigma(\alpha)$, so

$$v(\alpha') \ll \sigma(\theta \mid\!\sim \phi) M_\sigma(\alpha)^3$$

and the required result now follows as in the first case.

We have established that w satisfies K_σ. But this is impossible by Lemma 11 since $w(\beta) \gg M_\sigma(\beta)$, giving the required contradiction. $\blacksquare$

Proof of Theorem 5. Let $1 = \gamma_0$ and let $\gamma_1 \gg \gamma_2 \gg ...,\ \gamma_1, \gamma_2, ... \in B$, be the maximal representatives in B of each equivalence class (with respect to $\sim$) represented in B. Let $\lambda_i \in B$ be minimal such that $\lambda_i \sim \gamma_i$. Using Theorem 6 (whose proof does not depend on the result we are currently proving) let $\sigma : B \to K$ be a valid weighting such that $|\!\!\sim_{M_\sigma}$ is the rational closure of K. [Notice that according to our construction of the rational closure, $\mid\!\sim_{\vec{u}}$, $\vec{u}$ certainly is a vindicated solution of K.] Let $\rho : B \to K$ be a valid weighting such that amongst all such weightings the entropy of M_ρ is maximal. By the construction in the proof of Theorem 6 we may suppose that

$$\alpha \in u_i \iff M_\sigma(\alpha) \sim \gamma_i.$$

Let

$$s_i = \{\alpha \in At^L | M_\rho(\alpha) \sim \gamma_i\}.$$

It may appear that some of the s_i may be empty, indeed this will happen if there is an excess of γ's. Nevertheless if we temporarily drop that requirement of $\vec{s}$, whilst still retaining the same definition of $\mid\!\sim_{\vec{s}}$, then the proof of Theorem 1 shows that we shall still have $|\!\!\sim_{M_\rho} = \mid\!\sim_{\vec{s}}$.

We shall show that for as long as $u_j \neq \emptyset,\ u_j = s_j$, from which, using the construction of $\vec{u}$, the required result follows. So suppose not, say, $s_j = u_j$ for $j < i$ and $s_i \neq u_i$, so $s_i \subset u_i$ by the construction of $\vec{u}$. Let ν be an infinitesimal such that $\gamma_i \gg \nu \gg \gamma_{i+1}$. Define the probability function w to be the normalization of the function v where, for $\alpha \in \bigcup_j u_j$,

$$v(\alpha) = \begin{cases} M_\rho(\alpha)(1 + \lambda_j^2 \nu) & \text{if } M_\rho(\alpha) \sim \gamma_j,\ j \leq i, \\ \lambda_i^2 \nu & \text{if } \alpha \in u_i - s_i, \\ \lambda_j^2 \nu & \text{if } \alpha \in \bigcup_{j>i} u_j. \end{cases}$$

We shall now prove a couple of claims aimed at showing that there is a $\pi : K \to B$ such that $|\!\!\sim_{M_\pi}$ satisfies K and M_π has entropy exceeding that of M_ρ, and hence that our original assumption that $\vec{u} \neq \vec{s}$ must be false. The function π will be defined case-wise as we proceed through the claims.

Claim 1: Let $\theta \mid\!\sim \phi \in K$. If $\bigcup_{j<i} u_j \cap S_\theta \neq \emptyset$ set $\pi(\theta \mid\!\sim \phi)$ to be γ_i if $\rho(\theta \mid\!\sim \phi) \ll \gamma_i$ and $\rho(\theta \mid\!\sim \phi)$ otherwise. Then

$$w(\theta \wedge \neg\phi) \leq \pi(\theta \mid\!\sim \phi).w(\theta).$$

Proof Clearly it is enough to show this for v. Let j be minimal such that $u_j \cap S_\theta \neq \emptyset$, say $\alpha_0 \in u_j \cap S_\theta$, and suppose first that it is not the case that $\rho(\theta \mid\!\sim \phi) \ll \gamma_i$. Let $\min\{\gamma_{j+1}, \rho(\theta \mid\!\sim \phi)\} \sim \gamma_r$. Notice that $r \leq i$. Then

$$v(\theta \wedge \neg\phi) \leq M_\rho(\theta \wedge \neg\phi) + \text{terms in } \lambda_r^2\nu$$

whilst

$$v(\theta) \geq M_\rho(\theta) + \lambda_j^2\nu M_\rho(\alpha_0) - \text{terms } \lessapprox \lambda_{i+1}.$$

Hence it is enough to show that

$$M_\rho(\theta \wedge \neg\phi) + \text{terms in } \lambda_r^2\nu \leq \rho(\theta \mid\!\sim \phi)(M_\rho(\theta) + M_\rho(\alpha_0)\lambda_j^2\nu).$$

But this is clear since the inequality holds for the two leading M_ρ terms and $\rho(\theta \mid\!\sim \phi) \geq \lambda_r$ (by choice of λ_r) so $\lambda_r^2\nu < \rho(\theta \mid\!\sim \phi)\lambda_j^2\nu$.

Now assume instead that $\rho(\theta \mid\!\sim \phi) \ll \gamma_i$. Then $M_\rho(\theta \wedge \neg\phi) \ll \gamma_i$ so $S_{\theta\wedge\neg\phi} \cap \bigcup_{k\leq i} s_k = \emptyset$ and $v(\theta \wedge \neg\phi) \leq m\lambda_i^2\nu$ for some standard m. But

$$v(\theta) \geq M_\rho(\theta) \sim \gamma_j \gg \gamma_i$$

so

$$v(\theta \wedge \neg\phi) \leq \gamma_i v(\theta) \ll v(\theta)$$

as required.

Claim 2: Let $\theta \mid\!\sim \phi \in K$ and suppose that $j \geq i$ is such that $u_j \cap S_\theta \neq \emptyset$ whilst $u_k \cap S_\theta = \emptyset$ for $k < j$. Then

$$w(\theta \wedge \neg\phi) \leq \pi(\theta \mid\!\sim \phi)w(\theta) \ll w(\theta)$$

for $\pi(\theta \mid\!\sim \phi) = \gamma_{j+1}$.
Proof We have that $v(\theta) \geq \lambda_j^2\nu$ (even if $j = i$ and $(u_i - s_i) \cap S_\theta = \emptyset$ since then $M_\rho(\theta) \sim \gamma_i \gg \nu$) whilst $v(\theta \wedge \neg\phi) \leq m\lambda_{j+1}^2\nu$ for some standard m. The result follows.

Summing up we now have that for these $\pi : K \to B$ and w that

$$w(\theta \wedge \neg\phi) \leq \pi(\theta \mid\!\sim \phi).w(\theta)$$

for all $\theta \mid\!\sim \phi$ in K. Furthermore

$$\pi(\theta \mid\!\sim \phi).w(\theta) \ll w(\theta)$$

in all cases *except* possibly that given in Claim 1 when $\rho(\theta \mid\!\sim \phi) \gg \gamma_{j+1}$. It is also now clear, by a little elementary analysis, that the entropy of w exceeds that of M_ρ. In short this is because in the passage from M_ρ to w the scaling of the probabilities of the atoms in $\bigcup_{j\leq i} u_i$ produces (at worst) a reduction in entropy of magnitude $-c\nu \log(d)$ where $c, d \gg \nu$, the change of probabilities of the atoms in $\bigcup_{j>i} u_j$ produces (at worst) a reduction in entropy of magnitude $-mc\log(c)$ where $c \ll \nu$, m standard, whilst the the change of probabilities of the atoms in $u_i - s_i$ produces (at least) an increase in entropy of magnitude $-a\nu\log(\nu)$ for some $a \gg \nu$. This latter clearly dominates the terms of the form $-c\log(c)$ with $c \ll \nu$. It also dominates terms of the form $-c\nu\log(d)$ with $c, d \gg \nu$ since $-\log(\nu)$ is infinite with respect to $-(c/a)\log(d)$.

Since w satisfies K_π the entropy of M_π must also exceed that of M_ρ so in order to get the required contradiction all that remains is to show that $|\!\mid\!\sim_{M_\pi}$ satisfies K. That is, for each $\theta \mid\!\sim \phi$ in K we require that $M_\pi(\theta \wedge \neg\phi) \ll M_\pi(\theta)$. Since by Lemma 11 $M_\pi(\alpha) \gtrsim w(\alpha)$ for all atoms α it is clear that the requirement will hold whenever we have $\pi(\theta \mid\!\sim \phi).w(\theta) \ll w(\theta)$.

Thus the only point at which our requirement might conceivably fail is, as indicated above, when in Claim 1 $\rho(\theta \mid\!\sim \phi) \gg \gamma_{j+1}$. In this case we must show that for $\alpha \in S_{\theta\wedge\neg\phi}$ the constraints in K_π force $M_\pi(\alpha) \ll \gamma_j$. Now since $|\!\mid\!\sim_{M_\rho}$ does satisfy K for any such α there must, by Lemma 13, be a constraint $\xi \mid\!\sim \psi$ in K for which $\alpha \in S_{\xi\wedge\neg\psi}$ and $\rho(\xi \mid\!\sim \psi) \ll \gamma_j$. Let $\alpha \in u_r$. Obviously $r > j$ and $S_\xi \cap \bigcup_{k>r} u_k \neq \emptyset$. Let q be minimal such that $u_q \cap S_\xi \neq \emptyset$. If $q > i$ then $\pi(\xi \mid\!\sim \psi)$ will be one of γ_i or $\rho(\xi \mid\!\sim \psi)$ and both of these are infinitesimal with respect to γ_j, as required. On the other hand, if $q \leq i$ then perforce $\pi(\xi \mid\!\sim \psi) \lesssim \gamma_i$ so the required result holds again. ■

Proof of Theorem 6. The result from right to left follows from Lemma 13.

In the other direction suppose that $\vec{s} = s_0, s_1, ..., s_r$ is a vindicated solution of K and B is sufficiently large for K. Let $\gamma_0 = 1$ and $\gamma_1, \gamma_2, \ldots, \gamma_r \in B$ with

$$1 = \gamma_0 \gg \gamma_1 \gg \gamma_2 \gg ... \gg \gamma_r.$$

[Notice that since $\vec{s}$ is vindicated each $i = 1, 2, ..., r$ can be associated with a conditional in K, namely pick $\alpha \in s_i$ and associate i with that $\theta \mid\!\sim \phi \in K$ for which $\alpha \in S_{\theta\wedge\neg\phi}$ and $S_{\theta\wedge\neg\phi} \cap \bigcup_{j<i} s_j = \emptyset$, there are a

sufficient number of equivalence classes represented in B.]

Define v on the atoms by

$$v(\alpha) = \begin{cases} \gamma_i^2 & \text{if } \alpha \in s_i, \\ 0 & \text{if } \alpha \notin \bigcup_i s_i, \end{cases}$$

and $\sigma : K \to B$ by

$$\sigma(\theta \mid\!\sim \varphi) = \begin{cases} \gamma_i & \text{if i is least such that } s_i \cap S_{\theta\wedge\neg\varphi} \neq \emptyset, \\ \gamma_1 & \text{if } s_i \cap S_{\theta\wedge\neg\varphi} = \emptyset \text{ for all } i. \end{cases}$$

Let w be the ϵ-probability function corresponding to the normalization of v. It is easy to check that w, equivalently v, satisfies K_σ. For suppose $\theta \mid\!\sim \phi \in K$. If $S_{\theta\wedge\neg\phi} \cap \bigcup_i s_i \neq \emptyset$, say i is minimal such that $S_{\theta\wedge\neg\phi} \cap s_i \neq \emptyset$, then for some standard $m > 0$

$$v(\theta \wedge \neg\phi) \leq m\gamma_i^2 < \gamma_i\gamma_j^2 \leq \sigma(\theta \mid\!\sim \phi)v(\theta) \ll v(\theta),$$

where $j < i$ is minimal such that $S_\theta \cap s_j \neq \emptyset$. On the other hand if $S_{\theta\wedge\neg\phi} \cap \bigcup_i s_i = \emptyset$ then $v(\theta \wedge \neg\phi) = 0$ so v satisfies the corresponding constraint in K_σ trivially. Notice that this proof has, in addition, shown that $|\!\mid\!\sim_w$ satisfies K.

By Lemma 11 $M_\sigma(\alpha) \gtrsim w(\alpha)$ $(\sim v(\alpha))$ for all $\alpha \in At^L$ with $v(\alpha) \neq 0$. We now show that, conversely, $M_\sigma(\alpha) \gtrsim w(\alpha)$ for all $\alpha \in At^L$ with $M_\sigma(\alpha) \neq 0$. First notice that if $M_\sigma(\alpha) \neq 0$ then $\alpha \in \bigcup_i u_i$ (since $\mid\!\sim_{M_\sigma}$ is certainly a solution of K) so also $v(\alpha) \neq 0$. So suppose this holds. If $v(\alpha) = \gamma_i$, $i > 0$, then, because $\vec{s}$ is vindicated, i is minimal such that $S_{\theta\wedge\neg\phi} \cap s_i \neq \emptyset$ for some $\theta \mid\!\sim \phi \in K$ so $\sigma(\theta \mid\!\sim \phi) = \gamma_i$ and so $M_\sigma(\alpha) \leq \gamma_i \sim w(\alpha)$. On the other hand if $\alpha \in s_0$ then $\alpha \in u_0$ so by Lemma 11 $M_\sigma(\alpha) \sim 1 \sim w(\alpha)$. From this it follows that $|\!\mid\!\sim_{M_\sigma} = |\!\mid\!\sim_w = \mid\!\sim_{\vec{s}}$ and that these all satisfy K. ■

Proof of Theorem 7. Both parts of this theorem follow directly from Lemma 12 since Υ is definable in $\mathbb{R}^*$. ■

Before embarking on the prof of Theorem 8 we need to recall some notions and results from [2] and [3].

For K as usual a finite set of conditionals let

$$K^{\rightarrow} = \{\theta \to \varphi \mid \theta \mid\!\sim \phi \in K\}.$$

$K^{\rightarrow}$ is said to *exclude* $\psi \mid\!\sim \lambda$ if $K^{\rightarrow} \models \neg\psi$ and to *support* $\psi \mid\!\sim \lambda$ if $K^{\rightarrow} \not\models \neg\psi$ and $K^{\rightarrow} \models \psi \rightarrow \lambda$. Let $EX(K)$ denote the set of conditionals in K which K excludes.

Goodman and Bamber construct a directed graph $D(K)$ by taking the vertices of $D(K)$ to be all the subsets of K and the edges of $D(K)$ to be all ordered pairs (K_1, K_2) such that $K \supseteq K_1 \supset K_2 \supseteq EX(K_1)$.

To gain some insight into the structure of this graph let K_{se} be the result of repeatedly applying EX to K until no further contraction occurs. Then by noting that if $K \supseteq K' \supset K''$ then $EX(K') \supseteq EX(K'')$ we see that if $K \supseteq K' \supset K_{se}$ then $EX(K') \supseteq K_{se}$. Indeed $K' \supset EX(K')$ for there must be some k such that $EX^k(K) \supseteq K'$ but $EX^{k+1} \not\supseteq K'$, so $EX(K') \not\supseteq K'$. The conclusion we can draw from this is that every maximal path in this graph starts at K and ends at K_{se}.

The following result due to Bamber et al appears in [2] (as Theorem 2).

Lemma 14 *K strongly entails $\psi \mid\!\sim \lambda$ with near surety iff every maximal path in $D(K)$ contains a vertex which supports $\psi \mid\!\sim \lambda$ or $K_{se}^{\rightarrow} \models \neg\psi$.*

Proof of Theorem 8. We will show the finer result that for every maximal path $\mathcal{P}$ in $D(K)$ there is a vindicated solution $\mid\!\sim_{\vec{s}}$ of K for which

$$\psi \mid\!\sim_{\vec{s}} \lambda \iff \mathcal{P} \text{ supports } \psi \mid\!\sim \lambda \text{ or } K_{se} \models \neg\psi, \tag{4}$$

and conversely. We start in the right to left direction.

Let $K = K_0 \supset K_1 \supset \ldots \supset K_m = K_{se}$ be a maximal path in $D(K)$. Define $s_0, s_1, \ldots, s_{m-1}$ by

$$s_i = S_{\bigwedge K_i^{\rightarrow}} - \bigcup_{j<i} S_{\bigwedge K_j^{\rightarrow}}. \tag{5}$$

In this definition some of the s_i may be empty. However, as already remarked, we can still define $\mid\!\sim_{\vec{s}}$ in the usual way and this relation will not be altered by removing the copies of $\emptyset$ from $\vec{s}$. It will be clear from what follows that once these removals have been affected we shall have a vindicated solution of K satisfying the requirements of the theorem.

The first point to note (as already observed by Bamber et al), is that if this path is what Bamber et al refer to as *the path of steepest descent*, that is when $K_{i+1} = EX(K_i)$ for $i = 0, 1, ..., m-1$, then the resulting vector is precisely what we obtain by using the Lehmann-Magidor algorithm from [8] for the rational closure of K. In other words we obtain $\vec{u}$. It follows then that $u_0 = S_{\bigwedge K^{\rightarrow}}$ and

$$\bigcup_i u_i = S_{\bigwedge K^{\rightarrow}_{se}}$$

and hence that for our $\vec{s}$ as above $s_0 = u_0$ and

$$\bigcup_i s_i = \bigcup_i u_i.$$

The next step is to show that (4) holds for any conditional $\psi \mathrel{|\!\sim} \lambda$ where $\mathcal{P}$ is the path $K_0, K_1, ..., K_m$. First suppose $\psi \mathrel{|\!\sim}_{\vec{s}} \lambda$. If $s_i \cap S_\psi = \emptyset$ for all i then $S_{\bigwedge K^{\rightarrow}_{se}} \cap S_\psi = \emptyset$ so $K^{\rightarrow}_{se} \models \neg\psi$. Otherwise if i is minimal such that $s_i \cap S_\psi \neq \emptyset$ then $s_i \cap S_\psi \subseteq S_\lambda$. Hence $K^{\rightarrow}_i \not\models \neg\psi$ whilst $K^{\rightarrow}_j \models \neg\psi$ for $j < i$ so $K^{\rightarrow}_i \models \psi \rightarrow \lambda$, as required. Clearly the result in the other direction follows by just reversing this argument.

Since we now have that $\mathrel{|\!\sim}_{\vec{s}}$ satisfies K to complete this half of the proof it only remains to show that $\vec{s}$ is a vindicated solution, and the only step remaining there is to show that if $\alpha \in s_i$, $0 < i$, then there is $\theta \mathrel{|\!\sim} \phi \in K$ such that $\alpha \in S_{\theta\wedge\neg\phi}$ whilst $S_{\theta\wedge\neg\phi} \cap \bigcup_{j<i} s_j = \emptyset$. But in this case since $K_i \subset K_{i-1}$ there must be a $\theta \mathrel{|\!\sim} \phi \in K_{i-1}$ such that $\theta \mathrel{|\!\sim} \phi \notin K_i$ and $\alpha \notin S_{\theta\rightarrow\phi}$. Hence $\alpha \in S_{\theta\wedge\neg\phi}$ and $S_{\bigwedge K^{\rightarrow}_{i-1}} \subseteq S_{\theta\rightarrow\phi}$ so $s_j \subseteq S_{\theta\rightarrow\phi}$ for $j < i$, as required.

Turning now to the other direction assume that $\vec{s} = s_0, s_1, ..., s_m$ is a vindicated solution of K. For $i = 0, 1, ..., m$ let

$$K_i = \{\theta \mathrel{|\!\sim} \phi \in K \mid S_{\theta\wedge\neg\phi} \cap \bigcup_{j\leq i} s_j = \emptyset\}.$$

Let $\mathcal{P} = K_0, K_1, ..., K_m, EX(K_m)$ if $K_m \neq EX(K_m)$ and $K_0, K_1, ..., K_m$ otherwise. We shall show that $\mathcal{P}$ is a maximal path in $D(K)$ for which (4) holds.

Clearly $K_0 = K$ and for each $K_{i+1} \in \mathcal{P}$, $K_{i+1} \subseteq K_i$. Indeed this must be a strict inclusion because by vindication for any $\alpha \in s_{i+1}$ with $i < m$ there is some $\theta \mathrel{|\!\sim} \phi \in K$ with $\alpha \in S_{\theta\wedge\neg\phi}$ while $\bigcup_{j\leq i} s_j \cap S_{\theta\wedge\neg\phi} = \emptyset$, and so $\theta \mathrel{|\!\sim} \phi \in K_i - K_{i+1}$. Also since $\mathrel{|\!\sim}_{\vec{s}}$ satisfies K, $s_j \cap S_{\theta\wedge\neg\phi} = \emptyset$ whenever $\theta \mathrel{|\!\sim} \phi \in K_j$ for $j \leq i$ and so $s_j \subseteq S_{\bigwedge K^{\rightarrow}_j} \subseteq S_{\bigwedge K^{\rightarrow}_i}$ whenever

$j \le i$. Now if $\theta \mathrel{|\!\sim} \phi \in EX(K_i)$ then $K_i^{\rightarrow} \models \neg\theta$ so $s_j \subseteq S_{\neg\theta}$ for $j \le i$. This means that $S_\theta \cap \bigcup_{j \le i} s_j = \emptyset$ so since $\mathrel{|\!\sim}_{\vec{s}}$ satisfies K it is not possible that $s_{i+1} \cap S_{\theta\wedge\neg\phi} \neq \emptyset$. That is, it is not possible that $\theta \mathrel{|\!\sim} \phi \notin K_{i+1}$. Hence $K_{i+1} \supseteq EX(K_i)$.

Before showing that the last term in $\mathcal{P}$ is K_{se} we first show by induction on $i \le m$ that the identities (5) hold for these K_i. Clearly this holds for $i = 0$. Assume it holds below i. Then if $\alpha \in s_i$, $\alpha \notin S_{\theta\wedge\neg\phi}$ for each $\theta \mathrel{|\!\sim} \phi \in K_i$. Hence $\alpha \in S_{\theta\rightarrow\phi}$ for each $\theta \mathrel{|\!\sim} \phi \in K_i$, so $\alpha \in S_{\bigwedge K_i^{\rightarrow}}$. Also if $\alpha \in s_i$ then $\alpha \notin s_j$ for $j < i$ so, using the inductive hypothesis, $\alpha \notin \bigcup_{j<i} S_{\bigwedge K_j^{\rightarrow}}$.

In the other direction suppose $\alpha \in S_{\bigwedge K_i^{\rightarrow}} - \bigcup_{j<i} S_{\bigwedge K_j^{\rightarrow}}$. Then, by the inductive hypothesis, $\alpha \notin s_j$ for $j < i$. Also since $\alpha \in S_{\bigwedge K_i^{\rightarrow}} \subseteq S_{\bigwedge K_{se}^{\rightarrow}} = \bigcup_k s_k$, $\alpha \in s_k$ for some $k \ge i$. If $k > i$ then, since $\vec{s}$ is vindicated, there is a $\theta \mathrel{|\!\sim} \phi \in K$ such that $\alpha \in S_{\theta\wedge\neg\phi}$ and $\bigcup_{j<k} s_j \cap S_{\theta\wedge\neg\phi} = \emptyset$. But since $i < k$ this means that $\theta \mathrel{|\!\sim} \phi \in K_i$ so $\alpha \in S_{\theta\rightarrow\phi}$, contradiction.

Having established (5) we now show that the last term in $\mathcal{P}$ is K_{se}. If $K_m = EX(K_m)$ this is clear. Otherwise consider extending $\mathcal{P}$ to a maximal path

$$\mathcal{Q} = K_0, K_1, ..., K_m, EX(K_m), EX^2(K_m), ..., EX^q(K_m) = K_{se}.$$

Let $\vec{t}$ be defined from $\mathcal{Q}$ analogously to (5). Then $\vec{t}$ is a vindicated solution of K so $\bigcup_i t_i = \bigcup_i u_i = \bigcup_i s_i$ and for $i = 0, 1, ..., m$, $t_i = S_{\bigwedge K_i^{\rightarrow}} - \bigcup_{j<i} S_{\bigwedge K_j^{\rightarrow}} = s_i$. It follows then that $\emptyset = t_{m+1} = S_{\bigwedge EX(K_m)^{\rightarrow}} - S_{\bigwedge K_m^{\rightarrow}}$. Thus $S_{\bigwedge EX(K_m)^{\rightarrow}} = S_{\bigwedge K_m^{\rightarrow}}$ and

$$\begin{aligned} EX^2(K_m) &= \{\theta \mathrel{|\!\sim} \phi \in EX(K_m) \mid S_{\bigwedge EX(K_m)^{\rightarrow}} \models \neg\theta\ \} \\ &= \{\theta \mathrel{|\!\sim} \phi \in EX(K_m) \mid S_{\bigwedge K_m^{\rightarrow}} \models \neg\theta\ \} \\ &= EX(K_m). \end{aligned}$$

It follows that $EX(K_m) = K_{se}$, as required. To conclude the proof notice that (4) now follows as in the first half of the proof. ■

References

[1] Adams, E.W., *The Logic of Conditionals.* D.Reidel, Dordrecht, The Netherlands, 1975.

[2] Bamber, D and Goodman, I.R. Reasoning with Assertions of High Conditional Probability: Strong Entailment with Near Surety. *2nd*

International Symposium in Imprecise Probabilities and their Applications, Ithaca, New York 2001.

[3] Bamber, D. Entailment with Near Surety of Scaled Assertions of High Conditional Probability. *Journal of Philosophical Logic*, 29(1):1-74, 2000.

[4] Booth, R and Paris, J.B. A Note on the Rational Closure of Knowledge Bases with both Positive and Negative Knowledge.*Journal of Logic, Language and Information* 7(2): 165-190, 1998.

[5] Goldzsmidt, M, Morris,P and Pearl, J. A Maximum Entropy Approach to Nonmonotonic Reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(3):220-232, 1990.

[6] Hill, L.C. Doctorial Thesis, Manchester University, 2002.

[7] Kraus,S, Lehmann,D and Magidor,M. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence* 44(1-2):167-207, 1990.

[8] Lehmann,D and Magidor,M. What Does a Conditional Knowledge Base Entail?. *Artificial Intelligence* 55(1):1-60, 1992.

[9] Paris, J.B. *The Uncertain Reasoner's Companion. Cambridge Tracts in Theoretical Computer Science 39*, Cambridge University Press, 1994.

[10] Pearl, J., *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, San Mateo, CA, 1988.

[11] Shannon, C.F. A Mathematical Theory of Communication.*Bell System Tech. J.*, 27, 379-423 and 623-656, 1948.

[12] Shannon, C.F. and Weaver, W. *The Mathematical theory of Communication.*University of Illinois Press, Urbana, Illinois, 1949.

[13] van den Dries, L. Tame Topology and O-minimal Structures. *London Mathematical Society Lecture Note Series 248.* Cambridge University Press, 1998.

[14] Weydert, E. Defaults and Infinitesimals. Default Inference by Nonarchimedian Entropy Maximization. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI 95.* Morgan Kaufmann, 540-547, 1995.

[15] Wilkie, A. Model Completeness Results fo Expansions of th Real Field by Restricted Pfaffian Functions and the Exponential Function. *Journal of the American Mathematical Society* 9(4):1051-1054, 1996.