Pure Inductive Logic with Functions

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Pure Inductive Logic with Functions

E.Howarth[*] and J.B.Paris[†]

School of Mathematics

The University of Manchester

Manchester M13 9PL

lizhowarth@outlook.com, jeff.paris@manchester.ac.uk

June 13, 2017

Abstract

We consider the version of Pure Inductive Logic which obtains for the language with equality and a single unary function symbol giving a complete characterization of the probability functions on this language which satisfy Constant Exchangeability.

Key words: Constant Exchangeability, Inductive Logic, Logical Probability, Rationality, Uncertain Reasoning.

AMS Classification: 03B48, 03H10, 03A99.

Introduction

Conventional Inductive Logic as developed by Rudolf Carnap, see [1],[2],[3],[4], largely worked within the framework of a unary first order language and countably many constants, without equality or function symbols. With some few notable exceptions, in particular the paper [8] of Gaifman and Snir which this present paper may be seen as developing, most contributions to the area continued to use this rather impoverished context and it is only relatively recently that we have seen a concerted effort to understand the subject for more general languages with polyadic relations and equality, see for example [12], [14]. Adding function symbols to the language still remains somewhat neglected however, and it is to this topic that the present paper aims to make a contribution by giving a characterization along the lines of de Finetti's Theorem[1] of

1

those probability functions, for a language with just equality and a single unary function symbol, which satisfy the fundamental principle in Inductive Logic of Constant Exchangeability.[2]

There is, in our view, a good reason why adding function symbols to the language of Pure Inductive Logic is not entirely natural and from some standpoints this could explain why this field has largely lain fallow. To explain this reason we need to first explain what we see as the basic problem that Pure Inductive Logic is attempting to answer: Imagine an agent inhabiting a structure M for a predicate language L with all the elements of M named by the constant symbols $a_1, a_2, a_3, \ldots$ of L, and with the agent otherwise knowing nothing about M. We now ask the agent what probability she would give to some sentence θ of L being true in M.

Of course except in the situation when θ is a tautology or contradiction the agent would, on the face of it, have a completely free choice here. However we now add that the agent's answer should be *rational*, by which, for want of any better explanation, we mean that the agent should try to give the same answer as a second like-minded agent in a similar situation but with whom there can be no communication. One such immediate constraint that this imposes (we would claim) on our agent's answers as we vary θ is that they should be coherent as probabilities, that is that she is effectively picking a *probability function*[3] w on the set of sentences SL of L.

To date 'rationality' has been imposed or conferred on w by requiring that it satisfies certain, at least arguably rational, principles. For example, that it should respect symmetries, which may be justified in the way that we might argue it is rational in the absence of any further information to give a coin probability $1/2$ of landing heads. However a word of warning is needed here. In *Pure* Inductive Logic we assume that the agent in the unknown structure M genuinely has no knowledge about the ambient structure M; in particular the relation and constant symbols of L are, as far as she is concerned, devoid of any interpretation. It may well be that what one judges to be a 'rational principle' is based on some considerations of actual, interpreted, real world examples. Nevertheless, in Pure Inductive Logic it is the common idea behind these examples which is proposed as, to a greater or lesser extent, a rational principle.

Rather a wide assortment of such principles have been proposed in the usual situation where the language L has just relation and constant symbols, see for example [14], and surely there are many more still to come. However when it comes to adding equality, and even more so when adding functions, our intuitions[4] about what seems to be rational are currently far less well developed. One reason for this, we would suggest, is that the real world seems to provide but few examples of functions, and those it does provide are so context specific as to deny, or at least hide, any underlying rational principle.

Nevertheless in this paper we shall describe how Pure Inductive Logic may be extended to include the language with a single unary function symbol, the usual constants $a_1, a_2, \ldots$, and equality, but for the sake of simplicity no other relation symbols. The key aim of this paper will then be to prove a representation theorem for probability functions on this language which satisfy the sine qua non rational principle of Constant Exchangeability.

---

[2]Johnson's *Permutation Principle* (see [10]), Carnap's *Axiom of Symmetry* (see [15, page 974]), Kuiper's *Strong Principle of Order Indifference* (see [11]).

[3]To be defined shortly.

[4]At least these authors'.

Context

Apart from the concluding section we shall henceforth limit our attention to the first order language L with just equality, a single unary function symbol F and countably many constants a_n for $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$, the implicit intention being that these name every element of the universe. Let $SL/QFSL$ be the set of sentences/quantifier free sentences of L.

We define a *probability function* on SL to be a function $w : SL \to [0, 1]$ such that for all $\theta, \phi, \exists x \, \psi(x) \in SL$:

(P1) If $\vdash \theta$ then $w(\theta) = 1$.

(P2) If $\theta \vdash \neg\phi$ then $w(\theta \vee \phi) = w(\theta) + w(\phi)$.

(P3) $w(\exists x \, \psi(x)) = \lim_{n \to \infty} w(\bigvee_{i=1}^{n} \psi(a_i))$.

These (P1-3) are just as for the usual definition of a probability function in Pure Inductive Logic, see for example [14], when we do not have equality or function symbols, although now that we do the notion of logical implication in (P1-2) is enhanced accordingly to include the equality axioms[5] for L in addition to the usual axioms/rules of the Predicate Calculus. Notice that the so-called Gaifman's Axiom (P3) reflects the above mentioned intention that the a_n name all the elements in the universe.

For $b_1, b_2, \ldots, b_m$ distinct constants of L, $\Theta(b_1, b_2, \ldots, b_m) \in SL$ is a *state description* for $b_1, b_2, \ldots, b_m$ if it is consistent (with the axioms of equality) and of the form (up to logical equivalence)

$$\bigwedge_{i,j=1}^{m} (b_i = b_j)^{\epsilon_{i,j}} \wedge \bigwedge_{i,j=1}^{m} (F(b_i) = b_j)^{\delta_{i,j}}$$

where the $\epsilon, \delta \in \{0, 1\}$ and for $\phi \in SL$, $\phi^1 = \phi$, $\phi^0 = \neg\phi$. The upper case Greek letters $\Theta, \Upsilon, \Delta, \Phi, \Psi$ will always denote state descriptions.

In the study of Pure Inductive Logic when we do not have function symbols, state descriptions play an important role since by a result of Gaifman, [7], every probability function is determined simply by its values on state descriptions. Indeed if a probability function is defined on the quantifier free sentences to satisfy (P1-2) for these sentences then it extends uniquely to a probability function on all sentences. Once function symbols are added however that need no longer hold; we require also that already on the quantifier free sentences the probability function is forced to be total. In other words, for the case where we have a single function symbol F, for a unique extension to SL to exist we require (as originally observed by Gaifman & Snir [8, Basic Fact 1.2]) that already on $QFSL$,

(P4)   For all $i \in \mathbb{N}^+$,

$$\lim_{m \to \infty} w \big(\bigvee_{j=1}^{m} F(a_i) = a_j \big) = 1,$$

equivalently,

$$\lim_{n \to \infty} \lim_{m \to \infty} w \big(\bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} F(a_i) = a_j \big) = 1.$$

---

[5]As given for example in [13].

Notice that conversely if w is a probability function on SL, that is satisfies (P1-3), then it will satisfy (P4) since by (P3) this amounts to

$$w(\forall x \,\exists y \, F(x) = y) = 1$$

which, now that we have the equality axioms to hand, holds by (P1). [6]

Our representation theorem will assume the rationality assumption, almost universally accepted in this subject, that our probability functions are invariant under permutations of constants. More precisely, that they satisfy:

The Principle of Constant Exchangeability, Ex
If $\theta(a_1, a_2, \ldots, a_n) \in SL$ and $i_1, i_2, \ldots, i_n \in \mathbb{N}^+$ are distinct then

$$w(\theta(a_1, a_2, \ldots, a_n)) = w(\theta(a_{i_1}, a_{i_2}, \ldots, a_{i_n})).$$

It is straightforward to check that it suffices here to restrict θ to be a state description, exchangeability will then already be forced to hold for all of SL.

Throughout we shall assume, usually without explicit mention, that all the probability functions we are considering satisfy Ex. One useful consequence of this is that in general we do not need to be too fussy about the subscripts we are using for the a_i, just as long as they are different.

The Representation Theorem

In this section we will set up and explain our representation theorem of probability functions satisfying Ex on this language L with just equality and a single unary relation symbol F. The next section will then give a proof of this representation.

Before embarking on the technicalities it may be helpful to give an informal picture of what is going on. The main results of this paper, Theorems 1 and 2, say that any probability function on L satisfying Ex is actually a, possibly continuous, mixture of some very simple probability functions $v_{g,h}$, and conversely. So the basic pattern here resembles that of de Finetti's Theorem (in the notation of this paper see [14, page 55]) and several other representation theorems in the subject, see for example [12]. The functions $v_{g,h}$ also bear a similarity to some of the basic building block functions used in that latter paper, in particular the $u^{\bar{p}}$, in that we can think of them as randomly assigning *colours* to the constants according to some probabilities.

In more detail, we can think of $v_{g,h}$ as assigning colours from $\ldots, -2, -1, 0, 1, 2, \ldots$ (i.e. $\mathbb{Z}$) to the constants $a_1, a_2, a_3, \ldots$, the probability that a_i is assigned colour m being $g(m)$. The assignment of colours will uniquely determine the state descriptions that the constants satisfy. The equality relations which hold between the constants is determined by their given colours as follows: For $i \neq j$ if a_i, a_j both get assigned same colour $m > 0$ then it is taken that $a_i = a_j$. In all other cases it is taken that $a_i \neq a_j$. In particular then, even if a_i, a_j both get assigned the same colour $k \leq 0$ they are nevertheless still taken to be different, not equal.

---

[6]We do not need to also introduce additional constraints on w to ensure that the equality axioms get probability 1 in the extension since, their being Π_1, this follows from (P1) with $\theta \in QFSL$.

The function h specifies how the function F acts according to the colour of its argument. So suppose that i, j are distinct, a_i has been given colour $m \neq 0$ and a_j has been given colour k. If $h(m) = k$ then it is taken that $F(a_i) = a_j$, otherwise it is taken that $F(a_i) \neq a_j$. In the case however when a_i is assigned the colour 0 (usually thought of as uncompromisingly 'black') it is taken that $F(a_i) = a_i$ but $F(a_i) \neq a_j$ for $j \neq i$.

Overall, the probability the function $v_{g,h}$ gives to a state description $\Theta(a_1, \ldots, a_m)$ is the total probability of all ways of assigning colours to $a_1, \ldots, a_m$ consistent with h and such that Θ is the state description determined by that colouring.

It is convenient, and obviously nothing is lost, by having h defined only on colours with non-zero probability. Beyond that there are certain other properties which we need to impose on h. The first is that h only maps into those colours m with $m \geq 0$ and $g(m) > 0$, and $h(m) = 0$ just if $m = 0$ (and 0 is in the domain of h). Secondly we require that $h(-m) = m$ for $m \geq 0$ whenever $-m$ is in the domain of h. In turn this means that we have to require of g not only that the sum of the probabilities $g(i)$ for $i \in \mathbb{Z}$ is 1 but that if $g(-m) > 0$ and $m \geq 0$ then $g(m) > 0$.

We now restate this more formally.

Let $g : \mathbb{Z} \to \{r \in \mathbb{R} : r \geq 0\}$ and $h : S \to S$, where $S = \{m \in \mathbb{Z} : g(m) > 0\}$, be such that

$$
\begin{aligned}
&\text{(i) } \sum_{m \in \mathbb{Z}} g(m) = 1 \text{ and } g(n) \geq g(n+1) \text{ for } n \in \mathbb{N}^+. \\
&\text{(ii) } \text{If } m \in \mathbb{N} \text{ and } -m \in S \text{ then } m \in S \text{ and } h(-m) = m. \\
&\text{(iii) } \text{If } 0 < m \in S \text{ then } h(m) > 0. \tag{1}
\end{aligned}
$$

Notice then that if $0 \in S$ then $h(0) = 0$.

Let $\Theta(a_1, a_2, \ldots, a_m)$ be the state description

$$
\bigwedge_{i,j=1}^{m} (a_i = a_j)^{\epsilon_{i,j}} \wedge \bigwedge_{i,j=1}^{m} (F(a_i) = a_j)^{\delta_{i,j}}.
$$

We say that a colouring $\tau : \{1, \ldots, m\} \to S$ is *compatible* with Θ, h if for $1 \leq i, j \leq m$

$$
\tau(i) < 0 \text{ implies } ((\epsilon_{i,j} = 1 \iff i = j) \text{ and } (\delta_{i,j} = 1 \iff h(\tau(i)) = \tau(j))), \tag{2}
$$

$$
\tau(i) = 0 \text{ implies } (\epsilon_{i,j} = 1 \iff i = j \iff \delta_{i,j} = 1), \tag{3}
$$

$$
\tau(i), \tau(j) > 0 \text{ implies } ((\epsilon_{i,j} = 1 \iff \tau(i) = \tau(j)) \text{ and } (\delta_{i,j} = 1 \iff h(\tau(i)) = \tau(j))). \tag{4}
$$

Notice that in (2) $h(\tau(i)) = -\tau(i)$. Now set

$$
v_{g,h}(\Theta) = \sum_{\tau} \prod_{i=1}^{m} g(\tau(i)) \tag{5}
$$

where the sum is over those τ compatible with Θ, h.

Theorem 1. $v_{g,h}$ *extends to a probability function satisfying Ex.*

Proof. To show that $v_{g,h}$ extends to a probability function it is enough to check, see [14, page 42], that on state descriptions $\Theta(a_1, a_2, \ldots, a_m)$, $m \in \mathbb{N}$:

(i) $\quad v_{g,h}(\Theta(a_1, a_2, \ldots, a_m)) \geq 0,$

(ii) $\quad v_{g,h}(\top) = 1,$

(iii) $\quad v_{g,h}(\Theta(a_1, a_2, \ldots, a_m)) = \displaystyle\sum_{\Phi(a_1, \ldots, a_{m+1}) \models \Theta(a_1, \ldots, a_m)} v_{g,h}(\Phi(a_1, a_2, \ldots, a_{m+1})),$ (6)

and that the axioms of equality get probability 1. Of these (i) and (ii) are immediate. As for (iii) suppose that the state description $\Theta(a_1, a_2, \ldots, a_m)$ is

$$\bigwedge_{i,j=1}^{m} (a_i = a_j)^{\epsilon_{i,j}} \wedge \bigwedge_{i,j=1}^{m} (F(a_i) = a_j)^{\delta_{i,j}}$$

and τ is compatible with $\Theta(\vec{a}), h$, so τ makes a contribution $\prod_{i=1}^{m} g(\tau(i))$ to $v_{g,h}(\Theta(\vec{a}))$. Let $\nu : \{1, 2, \ldots, m+1\} \to S$ extend τ. Then to continue to satisfy (4) the choice of the additional $\epsilon_{s,t}, \delta_{s,t}$ for $s = m+1$ or $t = m+1$ is forced and there is a unique state description $\Phi(a_1, a_2, \ldots, a_{m+1})$, necessarily extending $\Theta(\vec{a})$, such that ν is compatible with $\Phi(a_1, a_2, \ldots, a_{m+1})$ and h. Since $\sum_{j \in \mathbb{Z}} g(j) = 1$,

$$\prod_{i=1}^{m} g(\tau(i)) = \prod_{i=1}^{m} g(\tau(i)) \cdot \left(\sum_{j \in \mathbb{Z}} g(j) \right) = \sum_{\nu} \prod_{i=1}^{m+1} g(\nu(i))$$

where the ν range over all extensions of τ to $\{1, 2, \ldots, m+1\}$, which gives, as required for (iii) of (6), that

$$v_{g,h}(\Theta(a_1, a_2, \ldots, a_m)) = \sum_{\Phi(\vec{a}, a_{m+1}) \models \Theta(\vec{a})} v_{g,h}(\Phi(\vec{a}, a_{m+1})).$$

To see that $v_{g,h}$ satisfies Ex notice that if σ is a permutation of $\{1, 2, \ldots, m\}$ and

$$\Psi(a_1, a_2, \ldots, a_m) \equiv \Theta(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)})$$

then τ is compatible with Ψ, h just if $\tau\sigma$ is compatible with $\Theta(a_1, a_2, \ldots, a_m), h$. Thus taking the sums over those τ, ν compatible with $\Psi(a_1, a_2, \ldots, a_m), \Theta(a_1, a_2, \ldots, a_m)$ and h respectively,

$$\sum_{\tau} \prod_{i=1}^{m} g(\tau(i)) = \sum_{\tau} \prod_{i=1}^{m} g(\tau(\sigma(i))) = \sum_{\nu} \prod_{i=1}^{m} g(\nu(i))$$

and

$$v_{g,h}(\Theta(a_1, a_2, \ldots, a_m)) = v_{g,h}(\Theta(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)}))$$

follows.

From this (or directly) it is easy to see that $v_{g,h}$ continues to satisfy its above definition given in terms of $a_1, a_2, \ldots, a_m$ if we replace these by any (distinct) constants for $b_1, b_2, \ldots, b_m$. Again we shall say that τ is compatible with $\Theta(b_1, b_2, \ldots, b_m), h$ if the above condition on compatibility holds with b_i in place of a_i.

6

Turning finally to (P4), it is enough by Ex to show that

$$\lim_{m \to \infty} v_{g,h}\left(\bigvee_{i=1}^{m} F(a_1) = a_i\right) = 1. \tag{7}$$

Consider a state description $\Phi(a_1, a_2, \ldots, a_m)$ such that

$$v_{g,h}\left(\Phi(\vec{a}) \wedge \neg \bigvee_{i=1}^{m} F(a_1) = a_i\right) > 0,$$

so since Φ is a state description $\Phi \models \neg \bigvee_{i=1}^{m} F(a_1) = a_i$. For $\tau : \{1, 2, \ldots, m\} \to S$ to be compatible with Φ, h it must be the case that $h(\tau(1)) \neq \tau(i)$ for $i = 1, 2, \ldots, m$. Thus summing over all state descriptions Φ for $a_1, a_2, \ldots, a_m$ we must have that

$$v_{g,h}\left(\neg \bigvee_{i=1}^{m} F(a_1) = a_i\right) \leq \sum_{h(\tau(1)) \notin Rg(\tau)} \prod_{i=1}^{m} g(\tau(i)).$$

But this sum is at most

$$\sum_{j \in S} g(j)(1 - g(h(j)))^{m-1}$$

which, since $g(h(j)) > 0$ for $j \in S$ and $\sum_j g(j) = 1$, tends to zero as $m \to \infty$. $\quad\square$

Notice that if $|\{i : g(i) > 0\}|$ is finite then for some m, $v_{g,h}$ will satisfy

$$v_{g,h}(\exists x_1, \ldots, x_m \, \forall y \, \bigvee_{i=1}^{m} y = x_i) = 1,$$

so giving probability 1 to the universe being finite. Conversely if $|\{i : g(i) > 0\}|$ is infinite then $v_{g,h}$ will give probability 0 to the universe being finite.

Having now shown that the $v_{g,h}$ are (after extension) probability functions on SL satisfying Ex we can state the main result of this paper:

The Representation Theorem 2. *Every probability function on SL satisfying Ex is a convex mixture of such $v_{g,h}$, and conversely.*

The proof of the converse here is straightforward (using for (P3) Lebesgue's Dominated Convergence Theorem) so we shall concentrate on the first part. This will be the content of the next section of this paper.

The Proof of the Representation Theorem

In this section Theorem 2 will be proved via several lemmas and a diversion into a nonstandard universe. Throughout, w will be a probability function on SL which satisfies Ex.

Lemma 3. *For any $\epsilon > 0$,*

$$\lim_{n \to \infty} w\left(\bigvee_{\Theta \in A_\epsilon(n)} \Theta(a_1, \ldots, a_n)\right) = 0,$$

where $A_\epsilon(n)$ is the set of state descriptions $\Theta(a_1, \ldots, a_n)$ for which

$$|\{i \leq n : \Theta(a_1, \ldots, a_n) \models \bigwedge_{j=1}^{n} F(a_i) \neq a_j\}| \geq \epsilon n.$$

Proof. Suppose on the contrary that this sequence was bounded below by $\gamma > 0$ for infinitely many n, for simplicity say all n. Notice that if $\Theta(a_1, \ldots, a_n) \in A_\epsilon(n)$ and σ is a permutation of $\{1, 2, \ldots, n\}$ then $\Theta(a_{\sigma(1)}, \ldots, a_{\sigma(n)}) \in A_\epsilon(n)$. Looking at the set of such $\Theta(a_{\sigma(1)}, \ldots, a_{\sigma(n)})$ as we vary σ we see that the proportion of $\Theta(a_{\sigma(1)}, \ldots, a_{\sigma(n)})$ from this set for which

$$\Theta(a_{\sigma(1)}, \ldots, a_{\sigma(n)}) \models \bigwedge_{j=1}^{n} F(a_1) \neq a_j$$

is at least ϵ. Hence, using Ex,

$$w\left(\bigvee\{\Theta(\vec{a}) \in A_\epsilon(n) : \Theta(\vec{a}) \models \bigwedge_{j=1}^{n} F(a_1) \neq a_j\}\right) \geq \epsilon\gamma.$$

So

$$\lim_{n \to \infty} w\left(\bigwedge_{i=1}^{n} F(a_1) \neq a_i\right) \neq 0$$

which contradicts that w satisfies (P4).

$\square$

Lemma 4. *Suppose that w is a probability function on SL, $r, m \in \mathbb{N}^+$ and ψ_i for $1 \leq i \leq rm\beta^{-1}$ are sentences of L with $w(\psi_i) \geq \beta > 0$. Then*

$$w\left(\bigvee_{\vec{\epsilon} : \sum_i \epsilon_i \geq m} \bigwedge_i \psi_i^{\epsilon_i}\right) \geq \beta(r-1)/r$$

where the $\epsilon_i \in \{0, 1\}$ for $1 \leq i \leq rm\beta^{-1}$.

Proof. Let

$$\Omega_{\vec{\epsilon}} = \bigwedge_i \psi_i^{\epsilon_i}, \qquad N(\vec{\epsilon}) = \sum_i \epsilon_i.$$

Notice that

$$\sum_{\vec{\epsilon}} N(\vec{\epsilon})\, w(\Omega_{\vec{\epsilon}}) = \sum_i w(\psi_i) \geq rm.$$

Let

$$B = \sum_{N(\vec{\epsilon}) < m} N(\vec{\epsilon})\, w(\Omega_{\vec{\epsilon}}), \quad C = \sum_{N(\vec{\epsilon}) \geq m} N(\vec{\epsilon})\, w(\Omega_{\vec{\epsilon}}).$$

Then

$$B < m \sum_{\vec{\epsilon}} w(\Omega_{\vec{\epsilon}}) = m,$$

so since $N(\vec{\epsilon}) \leq rm\beta^{-1}$,

$$w\big(\bigvee_{\vec{\epsilon}:\sum_i \epsilon_i \geq m} \bigwedge_i \psi_i^{\epsilon_i}\big) \geq \beta(rm)^{-1}C = \beta(rm)^{-1}\left(\sum_{\vec{\epsilon}} N(\vec{\epsilon})\, w(\Omega_{\vec{\epsilon}}) - B\right)$$
$$\geq \beta(rm)^{-1}(rm - m) = \beta(r-1)/r,$$

as required. $\qquad\square$

Notice that we cannot do much better than this because the ψ_i could all be the same sentence.

Lemma 5. *Suppose that $\phi(a_1, a_2) \in SL$ and*

$$w\big(\bigvee_{i=2}^{n+1} \phi(a_1, a_i)\big) > \beta > 0.$$

Then there is $\lambda > 0$ such that for all k eventually,[7]

$$w\big(|\{1 < i \leq k \,:\, \phi(a_1, a_i)\}| \geq \lambda k\big) > \beta. \tag{8}$$

Proof. Let

$$w\big(\bigvee_{i=2}^{n+1} \phi(a_1, a_i)\big) > \delta > \beta > 0.$$

Take ψ_j in Lemma 4 to be

$$\bigvee_{i=nj+2}^{n(j+1)+1} \phi(a_1, a_i),$$

and notice that by Ex $w(\psi_j) > \delta$. Furthermore in this case we actually have infinitely many of these ψ_j. By Lemma 4 we can now conclude that for any natural number m the probability w gives to the sentence

$$|\{1 \leq j \leq rm\beta^{-1} \,:\, \bigvee_{i=nj+2}^{n(j+1)+1} \phi(a_1, a_i)\}| \geq m$$

is at least $\delta(r-1)r^{-1}$ (since $\delta^{-1} < \beta^{-1}$). Now pick r so large that $\delta(r-1)r^{-1} > \beta$. Then for $r\beta^{-1} \leq k \in \mathbb{N}^+$,

$$w(|\{1 \leq j \leq k \,:\, \bigvee_{i=nj+2}^{n(j+1)+1} \phi(a_1, a_i)\}| \geq [kr^{-1}\beta]) > \beta,$$

where as usual $[..]$ indicates integer part of. Taking $\lambda = (2r)^{-1}\beta$ gives the result since $k(2r)^{-1}\beta < [kr^{-1}\beta]$ for all k eventually.

$\qquad\square$

Corollary 6. *Let $\beta > 0$. Then for some n,*

$$w\big(\bigvee_{i=2}^{n} \phi(a_1, a_i)\big) > \beta > 0 \tag{9}$$

---

[7]Notice that $|\{1 < i \leq k \,:\, \phi(a_1, a_i)\}| \geq \lambda k$ can be expressed as a sentence of L, and even a quantifier free sentence if ϕ is quantifier free.

just if there is $\lambda > 0$ such that for all k eventually,

$$w\big(|\{1 < i \leq k \,:\, \phi(a_1, a_i)\}| \geq \lambda k\big) > \beta. \tag{10}$$

Proof. The forward direction is just Lemma 5 (with n in place of $n+1$). In the other direction if (10) then for large k, $k > \lambda^{-1}$,

$$w\big(|\{1 < i \leq k \,:\, \phi(a_1, a_i)\}| \geq 1\big) > \beta,$$

which gives (9). $\qquad\qquad\square$

Let

$$\varsigma = \lim_{n \to \infty} w\big(\bigwedge_{i=2}^{n} F(a_1) \neq a_i\big).$$

Corollary 7. *For $0 < \beta < 1 - \varsigma$, there is a $\lambda > 0$ such that for all k eventually*

$$w\big(|\{1 < i \leq k \,:\, F(a_1) = a_i\}| \geq \lambda k\big) > \beta.$$

Proof. For large n,

$$w\big(\bigwedge_{i=2}^{n+1} F(a_1) \neq a_i\big) < 1 - \beta$$

so

$$w\big(\bigvee_{i=2}^{n+1} F(a_1) = a_i\big) > \beta.$$

The result now follows from Lemma 5 $\qquad\qquad\square$

The next result is not needed in the proof of Theorem 2 but maybe aids insight into what is going on.

Proposition 8.

$$\lim_{n \to \infty} w\big(F(a_1) = a_1 \wedge \bigwedge_{i=2}^{n} F(a_1) \neq a_i\big) = \varsigma.$$

Proof. From (P4) we must have that

$$\lim_{n \to \infty} w\big(\bigvee_{i=1}^{n} F(a_1) = a_i \wedge \bigwedge_{i=2}^{n} F(a_1) \neq a_i\big) = \varsigma$$

and the result follows. $\qquad\qquad\square$

Note that the probability of the sentence

$$|\{1 < i \leq k \,:\, F(a_1) = a_i\}| \geq \lambda k$$

is the same as the sum of the probabilities of state descriptions $\Theta(a_1, \ldots, a_k)$ which logically imply that sentence. As we shall see in future it will sometimes be useful to think of the probabilities of quantifier free sentences in terms of such sums.

By a directly similar proof to that of Lemma 5 but using

$$\phi(a_1, a_1) \vee \bigvee_{i=nj+2}^{n(j+1)+1} \phi(a_1, a_i)$$

10

throughout in place of

$$\bigvee_{i=nj+2}^{n(j+1)+1} \phi(a_1, a_i)$$

for $j = 0, 1, 2, \ldots$ we obtain:

Lemma 9. *Suppose that* $\phi(a_1, a_2) \in SL$ *and*

$$w\big(\phi(a_1, a_1) \vee \bigvee_{i=2}^{n+1} \phi(a_1, a_i)\big) > \beta > 0.$$

Then there is $\lambda > 0$ *such that for all* k *eventually,*

$$w\big(\phi(a_1, a_1) \vee |\{1 < i \leq k \,:\, \phi(a_1, a_i)\}| \geq \lambda k\big) > \beta.$$

Using this lemma, (P4) and $F(a_1) = a_2$ for $\phi(a_1, a_2)$ we now obtain:

Corollary 10. *For* $\beta < 1$, *there is a* $\lambda > 0$ *such that for all* k *eventually*

$$w\big(F(a_1) = a_1 \vee |\{1 < i \leq k \,:\, F(a_1) = a_i\}| \geq \lambda k\big) > \beta.$$

We now use the conventional trick in this area of moving to a nonstandard universe. Let U be a sufficiently large, transitive, set fragment of the 'standard' Set Theoretic Universe V in which all the objects required herein exist and have their usual properties. Let $U^* \in V$ be a countably saturated elementary extension of U (so with nonstandard natural numbers). In the proof that follows the reasoning will go backwards and forwards between V and U^*. Although whenever we feel there is danger of confusion we shall include a parenthetic clarification, still it seems expedient to say more about the relationship between V and U^* at this point.

Given $b \in U^* - U$ (the so called 'non-standard' elements) we will identify b with the set, in V,

$$\{c \in U^* \,:\, U^* \models c \in b\}. \tag{11}$$

Unfortunately we cannot in general continue to make this identification for $b \in U$ because, as in the case of $\mathbb{N}$, the set in (11) may acquire new non-standard elements not in b as such. For that reason for $b \in U$ we will denote the set in (11) by b^* *when this differs from* b. Note then that there will be no difference when $b \in V_{\omega_0}$ (so in particular for the individual natural numbers, codes for standard sentences etc.) so these will have the same denotation, b, throughout.

Let $\mathbb{N} < \nu \in \mathbb{N}^*$ and let $\mathbb{S}$ be the set (in U^*) of state descriptions $\Upsilon(a_1, \ldots, a_\nu)$.[8] Define (in the standard world V) a measure on the algebra of subsets of $\mathbb{S}$ (according to the above mentioned identification) which are in U^* by

$$X \mapsto {}^\circ\big(w^*\big(\bigvee X\big)\big),$$

where as usual $^\circ$ denotes the standard part, and by using Carathéodory's Theorem let μ be the σ-additive measure extending this measure on the σ-algebra (in the standard world V) generated by $\mathbb{S}$. (Notice that we can apply this theorem since if A_i for $i \in \mathbb{N}$

---

[8]Henceforth we will customarily shorten $\Upsilon(a_1, \ldots, a_\nu)$ to Υ.

are subsets of $\mathbb{S}$ in U^* and $\bigcup_{i \in \mathbb{N}} A_i = A \in U^*$ then for some $n \in \mathbb{N}$, $\bigcup_{i \leq n} A_i = A$, otherwise $\mathbb{N}$ would be definable in U^*.) Let

$$\mathbb{S}_0 = \{\Upsilon \in \mathbb{S} : w^*(\Upsilon) > 0\}.$$

Notice that $\mathbb{S}_0$ has μ-measure 1.

Lemma 11. *The set of $\Upsilon \in \mathbb{S}_0$ for which*

$$\nu^{-1}|\{i \leq \nu : \Upsilon \models \bigwedge_{j=1}^{\nu} F(a_i) \neq a_j\}| \text{ is infinitesimal (or 0)} \tag{12}$$

has μ-measure 1.[9]

Proof. From Lemma 3 for each $n \in \mathbb{N}^+$ w^* of the disjunction of the set of state descriptions $\Upsilon(a_1, \ldots, a_\nu) \in \mathbb{S}_0$ for which

$$|\{i \leq \nu : \Upsilon(a_1, \ldots, a_\nu) \models \bigwedge_{j=1}^{\nu} F(a_i) \neq a_j\}| \geq \nu/n$$

is infinitesimal. Hence this set has μ-measure 0. Taking the union over $n \in \mathbb{N}^+$ of these sets gives the result. $\qquad\square$

Let $\mathbb{S}_1$ be the measure 1 set of $\Upsilon \in \mathbb{S}_0$ satisfying (12) and let $\Gamma(\Upsilon)$ be the set of $\Upsilon(a_{\sigma(1)}, \ldots, a_{\sigma(\nu)})$ for σ a permutation of $\{1, \ldots, \nu\}$ (in U^*). Notice that if $\Upsilon \in \mathbb{S}_1$ then $\Gamma(\Upsilon) \subset \mathbb{S}_1$.

The objects which will be defined in what follows will generally be functions of Υ (or $\Gamma(\Upsilon)$). Sometimes these definitions, and the results that follow will only hold, or make sense, for a μ-measure 1 set of Υ. Although it will not always be convenient to make it explicit there will be an underlying assumption in these cases that the Υ concerned comes from this μ-measure 1 set. In view of the result we intend to prove, what happens outside of this μ-measure 1 set will be irrelevant.

For $\Upsilon \in \mathbb{S}_1$ set

$$i \sim_\Upsilon j \iff \Upsilon(a_1, \ldots, a_\nu) \models a_i = a_j$$

for $1 \leq i, j \leq \nu$. Let $H_1, \ldots, H_\xi$[10] be the equivalence classes of $\sim_\Upsilon$ and define a partial function ϱ on $\{1, \ldots, \xi\}$ to itself by setting $\varrho(i) = j$ if for each $k \in H_i$ there is $r \in H_j$ such that

$$\Upsilon \models F(a_k) = a_r.$$

Since $\Upsilon \in \mathbb{S}_1$

$$\nu^{-1}|\bigcup\{H_i : \varrho(i) \text{ not defined }\}| \text{ is infinitesimal.} \tag{13}$$

We may assume that the ordering here is chosen so that $H_1, \ldots, H_\kappa$ are the equivalences classes of $\sim_\Upsilon$ for which ϱ is defined and that

$$|H_1| \geq |H_2| \geq |H_3| \geq \ldots \geq |H_\kappa|.$$

---

[9]Henceforth we will take as read 'or 0' after 'infinitesimal'.

[10]These H_ζ, ξ etc. are functions of Υ but to simplify the notation we shall suppress explicitly indicating this dependence.

Notice that from (13) ϱ is defined on all the $m \in \mathbb{N}$ for which $\nu^{-1}|H_m|$ is not infinitesimal.

For $\Upsilon \in \mathbb{S}_1$ let

$$K_0 = \{i \,:\, \nu^{-1}|\{j \,:\, F(a_i) = a_j\}| \text{ is infinitesimal and non-zero}\},$$

where (as usual) we have abbreviated $\Upsilon \models F(a_i) = a_j$ to just $F(a_i) = a_j$. So K_0 is in the σ-algebra generated by the subsets of $\{1, 2, \ldots, \nu\}$ in U^*. Similarly the set of Υ for which $1 \in K_0$ is in the domain of μ and has measure ς. This follows since[11]

$$
\begin{aligned}
\mu(1 \in K_0) &= \lim_{n \to \infty} \mu\big(0 < \nu^{-1}|\{j \,:\, F(a_1) = a_j\}| \leq n^{-1}\big) \\
&= \lim_{n \to \infty} \mu\big(\nu^{-1}|\{j \,:\, F(a_1) = a_j\}| \leq n^{-1}\big) \\
&= 1 - \lim_{n \to \infty} \mu\big(\nu^{-1}|\{j \,:\, F(a_1) = a_j\}| > n^{-1}\big) \\
&= 1 - \lim_{n \to \infty} \mu\big(\bigvee_{i=2}^{n} F(a_1) = a_i\big), \\
&= \lim_{n \to \infty} \mu\big(\bigwedge_{i=2}^{n} F(a_1) \neq a_i\big) \\
&= \varsigma.
\end{aligned}
$$

Here the second equality follows from Lemma 11 (and arguing as in Lemma 3 to get the property for a_1) and the fourth by Corollary 6 and overspill (since U^* is an elementary extension of U).

Notice that for all $\Upsilon \in \mathbb{S}_1$ for which $1 \in K_0$ then there is no (standard) $\lambda > 0$ such that

$$|\{j \,:\, F(a_1) = a_j\}| > \lambda\nu.$$

So for each $n \in \mathbb{N}$,

$$\mu\big(1 \in K_0 \wedge \bigvee_{i=2}^{n} F(a_1) = a_i\big) = 0.$$

Hence by (P4) we must have that

$$\mu(1 \in K_0 \wedge F(a_1) = a_1) = \mu(1 \in K_0) \tag{14}$$

Notice that (for μ-measure 1 state descriptions Υ, as usual) if $i \in K_0$ and $i \in H_\varsigma$ then $F(a_i) = a_i$, $\varrho(\zeta) = \zeta$ and $\nu^{-1}|\{j \,:\, F(a_i) = a_j\}|$ is infinitesimal so $\nu^{-1}|H_\zeta|$ is infinitesimal.

Lemma 12. *For a μ-measure 1 set of $\Upsilon \in \mathbb{S}_1$ either $F(a_1) = a_1$ or*

$$^{\circ}(\nu^{-1}|\{i \leq \nu \,:\, F(a_1) = a_i\}|) > 0.$$

Proof. From Corollary 10 for each $n \in \mathbb{N}^+$ there is a $0 < \lambda \in \mathbb{R}$ such that for all k eventually (in V)

$$\mu\big(F(a_1) = a_1 \vee k^{-1}|\{1 < i \leq k \,:\, F(a_1) = a_i\}| \geq \lambda\}\big) \geq 1 - n^{-1}.$$

---

[11]By $\mu(1 \in K_0)$ we mean μ of the set of Υ for which $1 \in K_0$.

Hence the set of $\Upsilon \in \mathbb{S}_1$ such that either $\Upsilon \models F(a_1) = a_1$ or

$$^\circ(\nu^{-1}|\{i \,:\, \Upsilon \models F(a_1) = a_i\}|) > 0$$

has μ-measure at least $1 - n^{-1}$ for each $n \in \mathbb{N}^+$. The result follows.

$\square$

Lemma 13. *Let* $0 < \epsilon \le 1$. *Then for all* $n \in \mathbb{N}$ *eventually*

$$\mu\{\Upsilon \,:\, |\{j \,:\, F(a_j) = a_j \vee |\{i \,:\, F(a_j) = a_i\}| > \nu/n\}| > (1 - \epsilon)\nu\} > 1 - \epsilon.$$

Proof. By Lemma 12, for large n

$$\mu\{\Upsilon \,:\, F(a_1) = a_1 \vee |\{i \,:\, F(a_1) = a_i\}| > \nu/n\} > 1 - \epsilon^2. \qquad (15)$$

By Ex (15) holds too for any a_j in place of a_1. Let $\Upsilon(j) = 1$ if

$$\Upsilon \models F(a_j) = a_j \vee |\{i \,:\, F(a_j) = a_i\}| > \nu/n$$

and 0 otherwise. Then from (15),

$$\sum_\Upsilon w^*(\Upsilon) \cdot \Upsilon(j) > 1 - \epsilon^2.$$

Summing over j and reversing the order of summation gives

$$\sum_\Upsilon w^*(\Upsilon) \sum_j \Upsilon(j) > \nu(1 - \epsilon^2). \qquad (16)$$

Now let

$$M = \sum\{w^*(\Upsilon) \,:\, \sum_j \Upsilon(j) \le \nu(1 - \epsilon)\},$$

so

$$\sum\{w^*(\Upsilon) \,:\, |\{j \,:\, F(a_j) = a_j \vee |\{i \,:\, F(a_j) = a_i\}| > \nu/n\}|$$
$$= \sum\{w^*(\Upsilon) \,:\, \sum_j \Upsilon(j) > \nu(1 - \epsilon)\} \ge 1 - M. \qquad (17)$$

Over estimating the sum $\sum_j \Upsilon(j)$ if it is less or equal $\nu(1 - \epsilon)$ by $\nu(1 - \epsilon)$ and by ν otherwise gives with (16),

$$\nu(1 - \epsilon^2) < M\nu(1 - \epsilon) + (1 - M)\nu$$

which simplifies to $M < \epsilon$. Hence, with (17), as required,

$$\sum\{w^*(\Upsilon) \,:\, |\{j \,:\, F(a_j) = a_j \vee |\{i \,:\, F(a_j) = a_i\}| > \nu/n\}| > \nu(1 - \epsilon)\} > 1 - \epsilon.$$

$\square$

Corollary 14. *For* $k \in \mathbb{N}^+$ *and a* μ-*measure* 1 *set of* $\Upsilon \in \mathbb{S}_1$ *either* $^\circ(\nu^{-1}|H_k|) = 0$ *or* $^\circ(\nu^{-1}|H_{\varrho(k)}|) > 0$.

14

Proof. Suppose on the contrary that this fails, so for some k there is a μ-measure $\beta > 0$ set of Υ,

$$^\circ(\nu^{-1}|H_k|) > 0 \text{ and } ^\circ(\nu^{-1}|H_{\varrho(k)}|) = 0. \tag{18}$$

Let $m \in \mathbb{N}$ be such that for a set of Υ of μ-measure at least $\beta/2$,

$$\nu^{-1}|H_k| > m^{-1} \text{ and } ^\circ(\nu^{-1}|H_{\varrho(k)}|) = 0. \tag{19}$$

Let $\epsilon = \min\{^\circ(\nu^{-1}|H_k|)/2, \beta/4\}$. By Lemma 13 there is an $n > m$ such that for a set of Υ of μ-measure at least $1 - \epsilon$,

$$|\{j \,:\, F(a_j) = a_j \vee \nu^{-1}|\{i \,:\, F(a_j) = a_i\}| > 1/n\}| > (1 - \epsilon)\nu. \tag{20}$$

Since $\beta/2 > \epsilon$ there is a set X of Υ of μ-measure at least $\beta/2 - \epsilon > 0$ for which (19) and (20) both hold. Since $\nu^{-1}|H_k| > \epsilon$ for $\Upsilon \in X$ there must be $r \in H_k$ such that

$$r \in \{j \,:\, F(a_j) = a_j \vee \nu^{-1}|\{i \,:\, F(a_j) = a_i\}| > 1/n\}.$$

Since for such Υ we have that $^\circ(\nu^{-1}|H_{\varrho(k)}|) = 0$ the disjunct

$$\nu^{-1}|\{i \,:\, F(a_j) = a_i\}| > 1/n$$

must fail for $j = r$, so that $F(a_r) = a_r$. But then $\varrho(k) = k$ and

$$^\circ(\nu^{-1}|H_{\varrho(k)}|) = ^\circ(\nu^{-1}|H_k|) > 0,$$

contradiction.

$\square$

For $\Upsilon \in \mathbb{S}_1$, $n \in \mathbb{N}^+$, $\zeta \leq \kappa$ set[12]:

$$Y = \bigcup\{H_\beta \,:\, \varrho(\beta) \text{ not defined}\},$$

$$K_{0,n} = \bigcup\{H_\beta \,:\, \nu^{-1}|H_\beta| < n^{-1}, \varrho(\beta) \text{ defined and } \nu^{-1}|H_{\varrho(\beta)}| < n^{-1}\},$$

$$K_{\zeta,n} = \bigcup\{H_\beta \,:\, \nu^{-1}|H_\beta| < n^{-1}, \varrho(\beta) = \zeta \text{ defined and } \nu^{-1}|H_{\varrho(\beta)}| \geq n^{-1}\},$$

$$B_{\zeta,n} = \begin{cases} H_\zeta & \text{if } \nu^{-1}|H_\zeta| \geq n^{-1}, \\ \emptyset & \text{otherwise.} \end{cases}$$

Note that for fixed n these sets $K_{0,n}, K_{\zeta,n}, B_{\zeta,n}, Y$ are all in U^* and that together they form a partition of $\{1, 2, \ldots, \nu\}$ since for $\Upsilon \in \mathbb{S}_1$ (so $\nu^{-1}|Y|$ is infinitesimal) and $^\circ(\nu^{-1}|H_\zeta|) > 0$, $F(a_r)$ must be defined on $r \in H_\zeta$ and so $\varrho(\zeta)$ will be defined. Hence for $n \in \mathbb{N}^+$ (and a μ-measure 1 set of Υ), in U^*,

$$|K_{0,n}| + \sum_\zeta |K_{\zeta,n}| + \sum_\zeta |B_{\zeta,n}| + |Y| = \nu. \tag{21}$$

Furthermore, using Corollary 14,

$$K_0 = \bigcap_{n \in \mathbb{N}^+} K_{0,n}.$$

___

[12]Again these $Y, K_{0,n}$ etc. are functions of Υ but to simplify the notation we shall suppress explicitly indicating this dependence.

Let

$$K_\zeta = \bigcup_{k \in \mathbb{N}^+} \bigcap_{k \le n \in \mathbb{N}^+} K_{\zeta,n}$$

$$B_\zeta = \bigcup_{n \in \mathbb{N}^+} B_{\zeta,n},$$

so B_ζ is H_ζ if $\nu^{-1}|H_\zeta|$ is not infinitesimal, otherwise it is $\emptyset$. Clearly then $B_\zeta = \emptyset$ for $\mathbb{N} < \zeta$.

Also $K_\zeta = \emptyset$ for $\mathbb{N} < \zeta$. For suppose on the contrary that $\emptyset \ne H_\beta \subset K_\zeta$. Then $\nu^{-1}|H_\beta|$ is infinitesimal whilst $\varrho(\beta) = \zeta$ and $\nu^{-1}|H_{\varrho(\beta)}| \ge n^{-1}$ for some $n \in \mathbb{N}^+$. But since the sizes of the H_γ are decreasing this means $\zeta \le n$ (contradiction) since if $n < \zeta$ then $H_1, H_2, \ldots, H_{n+1}$ would be disjoint subsets of $\{1, 2, \ldots, \nu\}$ each of size at least ν/n, which is impossible.

By the same argument, $B_{\zeta,n} = K_{\zeta,n} = \emptyset$ for $\zeta > n$.

To sum up then we now have that (for a μ-measure 1 set) all except an infinitesimal fraction of $i \in \{1, 2, \ldots, \nu\}$ are in (exactly) one of

$$B_1, B_2, \ldots, B_n, \ldots, \quad K_0, K_1, K_2, \ldots, K_n, \ldots \text{ for } n \in \mathbb{N}.$$

Now for $\Upsilon \in \mathbb{S}$, $\Gamma = \Gamma(\Upsilon)$, set, as far as possible,[13]

$$g(0) = \lim_{n \to \infty} {}^\circ(\nu^{-1}|K_{0,n}|) \text{ and } h(0) = 0 \text{ when } g(0) > 0,$$

and for $m \in \mathbb{N}^+$

$$g(m) = {}^\circ(\nu^{-1}|B_m|) \text{ and } h(m) = k \text{ when } \varrho(m) = k \text{ and } g(m) > 0,$$

and

$$g(-m) = \lim_{n \to \infty} {}^\circ(\nu^{-1}|K_{m,n}|) \text{ and } h(-m) = m \text{ if } g(-m) > 0.$$

Since they are monotone the sequences for $g(m)$ with $m \ge 0$ do converge. As for the case of $g(-m)$ when $m > 0$ notice that for $r \ge n$,

$$\left||K_{m,n}| - |K_{m,r}|\right| \le |K_{0,n}| - |K_{0,r}| + \sum\{|H_\beta| : r^{-1} \le \nu^{-1}|H_\beta| < n^{-1}\}$$

since the only H_β which are subsets of just one of $K_{m,n}, K_{m,r}$ must either have been a subset of $K_{0,n}$ but not $K_{0,r}$ or must have become one of the $B_{\beta,j}$ for some $n \le j \le r$. Hence the sequence for $g(-m)$ also converges.

Notice that these definitions do indeed only depend on $\Gamma = \Gamma(\Upsilon)$.

The plan is now to show that on standard state descriptions, $v_{g,h}$ is infinitesimally close to w^* conditioned on $\bigvee \Gamma$. First however we need to show that g, h satisfy the conditions specified in the definition of $v_{g,h}$. These are the contents of the next three lemmas.

Lemma 15.
$$\sum_{i \in \mathbb{Z}} g(i) = 1 \text{ and } g(n) \ge g(n+1) \text{ for } n \in \mathbb{N}^+$$

for a μ-measure 1 set of Γ.

---

[13]Again g, h are functions of Γ but we suppress explicit mention of this argument.

Proof. The second part is immediate since the $|H_\beta|$ are decreasing in β.

Turning to the first part, from (21) and the fact that $B_{\zeta,n} = K_{\zeta,n} = \emptyset$ for $\zeta > n$ we obtain that

$$|K_{0,n}| + \sum_{\zeta \le n} |K_{\zeta,n}| + \sum_{\zeta \le n} |B_{\zeta,n}| + |Y| = \nu.$$

Hence taking standard parts

$$^\circ(\nu^{-1}|K_{0,n}|) + \sum_{\zeta \le n} {}^\circ(\nu^{-1}|K_{\zeta,n}|) + \sum_{\zeta \le n} {}^\circ(\nu^{-1}|B_{\zeta,n}|) = 1, \tag{22}$$

since $\nu^{-1}|Y|$ is infinitesimal. Now given $\epsilon > 0$ let $\epsilon^{-1} \le m_1 \in \mathbb{N}$ be such that $^\circ(\nu^{-1}|K_{0,n}|)$ is within ϵ of $g(0)$ for $n \ge m_1$ and let $m_1 \le m_2 \in \mathbb{N}$ be such that for $0 < \zeta < m_1$ and $n \ge m_2$ each of the $^\circ(\nu^{-1}|K_{\zeta,n}|)$ are within $m_1^{-1}\epsilon$ of $g(-\zeta)$. Then by considering how the H_β move between the $K_{0,n}, K_{\zeta,n}, B_{\zeta,n}$ we see that

$$^\circ(\nu^{-1}|K_{0,n}|) + \sum_{\zeta \le m_2} {}^\circ(\nu^{-1}|K_{\zeta,n}|) + \sum_{\zeta \le m_2} {}^\circ(\nu^{-1}|B_{\zeta,n}|)$$

must be at least $1 - 2\epsilon$ for all $n \ge m_2$. Hence

$$\sum_{i=-m_2}^{m_2} g(i) \ge 1 - 2\epsilon$$

and the required result follows.

$\square$

Lemma 16. *Let $m \in \mathbb{N}^+$. Then for a μ-measure 1 set of state descriptions $\Upsilon \in \mathbb{S}_1$, if $g(-m) > 0$ then $g(m) > 0$.*

Proof. Suppose that $g(-m) > 0$. Then certainly for some n we must have that $|K_{m,n}| > 0$. But that means that for some β, $H_\beta \subseteq K_{m,n}$, $\varrho(\beta) = m$ and $\nu^{-1}|H_{\varrho(\beta)}| \ge n^{-1}$. Hence

$$g(m) = {}^\circ(\nu^{-1}|B_m|) = {}^\circ(\nu^{-1}|H_m|) \ge n^{-1} > 0.$$

$\square$

Lemma 17. *Let $m \in \mathbb{N}^+$. Then for a μ-measure 1 set of state descriptions $\Upsilon \in \mathbb{S}_1$, if $g(m) > 0$ then $h(m)$ is defined, $h(m) > 0$ and $g(h(m)) > 0$.*

Proof. For $g(m) > 0$ with $m \in \mathbb{N}^+$ we must have that $^\circ(\nu^{-1}|H_m|) > 0$. The result now follows from Corollary 14. $\square$

To summarize at this point, we have a μ-measure 1 set $\mathbb{S}_1$ of state descriptions, alternately sets $\Gamma(\Upsilon)$ since if $\Upsilon \in \mathbb{S}_1$ and $\Delta \in \Gamma(\Upsilon)$ then $\Delta \in \mathbb{S}_1$. For each such Γ we have defined functions $g : \mathbb{Z} \to \{r \in \mathbb{R} : r \ge 0\}$ and $h : S \to S$, where $S = \{m \in \mathbb{Z} : g(m) > 0\}$, such that

(i) $\displaystyle\sum_{m \in \mathbb{Z}} g(m) = 1$ and $g(m) \ge g(m+1)$ for $m \in \mathbb{N}^+$.

(ii) If $m \in \mathbb{N}$ and $-m \in S$ then $m \in S$ and $h(-m) = m$.

(iii) If $0 < m \in S$ then $h(m) > 0$.

So these g, h, S satisfy the requirements given in (1) and $v_{g,h}$ is defined.

17

Lemma 18. *Let $\Theta(a_1, \ldots, a_m)$ be a standard (i.e. in V) state description. Then for a μ-measure 1 set of $\Upsilon \in \mathbb{S}_1$ (or $\Gamma = \Gamma(\Upsilon)$),*[14]

$$| \, v_{g,h}(\Theta) - w^*(\Theta \, | \, \bigvee \Gamma) \, | \tag{23}$$

is infinitesimal.

Proof. We first derive arbitrarily close approximations to the terms in (23). Starting with the right hand term notice that

$$
\begin{aligned}
w^*(\Theta \wedge \bigvee \Gamma) \;&=\; \sum_{\Delta \in \Gamma} w^*(\Theta \wedge \Delta) \\
&=\; \sum \{ w^*(\Delta) \, : \, \Delta \in \Gamma \text{ and } \Delta \models \Theta \} \\
&=\; w^*(\Upsilon) \cdot |\{ \Delta \in \Gamma \, : \, \Delta \models \Theta \}|, \text{ since } w^* \text{ satisfies Ex.}
\end{aligned}
$$

Now $|\{ \Delta \in \Gamma \, : \, \Delta \models \Theta \}|$ is just the number of permutations σ of ν (in U^*) divided by the number of these permutations which fix Υ (that is, the number of elements in Γ) all multiplied by the proportion of these for which

$$\Upsilon(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(\nu)}) \models \Theta(a_1, a_2, \ldots, a_m), \tag{24}$$

equivalently multiplied by the probability (in U^*) of picking *without replacement* $\sigma(1)$, $\sigma(2)$, $\ldots, \sigma(m)$ from $\{1, 2, \ldots, \nu\}$ such that

$$\Upsilon(a_1, a_2, \ldots, a_\nu) \models \Theta(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)}). \tag{25}$$

Since $\nu > \mathbb{N}$ this probability will differ by at most an infinitesimal if we drop the requirement of not replacing previous choices. So if we take standard $\epsilon \in (0, 1/2)$ we can from now on assume, with a difference in overall probability of at most ϵ, that the choice is made with replacement. This probability of picking $\sigma(1), \ldots, \sigma(m)$ from $\{1, 2, \ldots, \nu\}$ with replacement such that (25) holds will be our starting approximation $\mathcal{A}_R$ to the right hand side of (23).

Turning now to the left hand term in (23) let $m_\epsilon \in \mathbb{N}^+$ be large enough that $m_\epsilon \geq \epsilon^{-1}$, the statement of Lemma 13 holds with $n = m_\epsilon$, and

$$\sum_{|i| > m_\epsilon} g(i) < \epsilon/m.[15] \tag{26}$$

We now use m_ϵ to approximate the expression (5) which gives the left hand side of (23). Let $j \in \mathbb{Z}$ and $1 \leq k \leq m$. Then if we change the sum in (5) to *all* $\tau : \{1, 2, \ldots, m\} \to S$ such that $\tau(k) = j$ we obtain

$$\sum_\tau \prod_{i=1}^m g(\tau(i)) = g(j) \sum_\tau \prod_{i \neq k} g(\tau(i)) = g(j) \prod_{i \neq k} \left(\sum_{r \in \mathbb{Z}} g(r) \right) = g(j).$$

---

[14] We use the standard notation here that $\bigvee \Gamma$ stands for $\bigvee_{\Delta \in \Gamma} \Delta(a_1, \ldots, a_\nu)$

[15] Note that only by this latter requirement does m_ϵ become dependent on Υ (for $\Upsilon \in \mathbb{S}_1$) and in fact that too can also be avoided by restricting to, say, a μ-measure $1 - \epsilon$ subset of these Υ.

It follows that the sum of the contributions $\sum_\tau \prod_{i=1}^m g(\tau(i))$ to $v_{g,h}(\Theta)$ where τ is compatible with Θ, h and some $\tau(i)$ is outside $[-m_\epsilon, m_\epsilon]$ is at most ϵ. The reduced sum for (5), with the range of τ restricted to $[-m_\epsilon, m_\epsilon]$, will be our starting approximation $\mathcal{A}_L$ to the left hand side of (23).

Now let $m_\epsilon \le r_\epsilon \in \mathbb{N}$ be *large*[16] and such that

$$| g(0) - \nu^{-1}|K_{0,r_\epsilon}| | + \sum_{j=1}^{m_\epsilon} | g(-j) - \nu^{-1}|K_{j,r_\epsilon}| | + \sum_{j=1}^{m_\epsilon} | g(j) - \nu^{-1}|B_{j,r_\epsilon}| | < \frac{\epsilon}{m}. \quad (27)$$

Let $\tau : \{1, 2, \ldots, m\} \to [-m_\epsilon, m_\epsilon]$ (so if τ is compatible with Θ, h it gives one of the summands in $\mathcal{A}_L$) and let

$$X_i = \begin{cases} B_{j,r_\epsilon} & \text{if } \tau(i) = j \in [1, m_\epsilon], \\ K_{0,r_\epsilon} & \text{if } \tau(i) = 0, \\ K_{j,r_\epsilon} & \text{if } \tau(i) = -j \in [-m_\epsilon, -1]. \end{cases} \quad (28)$$

(Notice that for $j \in X_i$, $F(a_j)$ is defined.) Then the probability of picking $\sigma(1), \sigma(2), \ldots, \sigma(m)$ so that for $i = 1, 2, \ldots, m$, $\sigma(i) \in X_i$ is $\prod_{i=1}^m \varkappa_i$, where $\varkappa_i = \nu^{-1}|X_i|$, and

$$\begin{aligned} | \prod_{i=1}^m \varkappa_i - \prod_{i=1}^m g(\tau(i)) | &= | \sum_{j=1}^m (\prod_{i<j} \varkappa_i)(\varkappa_j - g(\tau(j)))(\prod_{j<i\le m} g(\tau(i))) | \\ &\le | \sum_{j=1}^m (\prod_{i<j} \varkappa_i)(|\varkappa_j - g(\tau(j))|)(\prod_{j<i\le m} g(\tau(i))) | \\ &\le \frac{\epsilon}{m} \sum_{j=1}^m (\prod_{i<j} \varkappa_i)(\prod_{j<i\le m} g(\tau(i))). \quad (29) \end{aligned}$$

by (27).

Since

$$\sum_{j=-m_\epsilon}^{m_\epsilon} g(j), \quad \sum_{j=1}^{m_\epsilon} \nu^{-1}|B_{j,r_\epsilon}| + \sum_{j=-m_\epsilon}^{0} \nu^{-1}|K_{j,r_\epsilon}|$$

are both at most 1,

$$\mathcal{A}_C = \sum_\tau \prod_{i=1}^m \varkappa_i,$$

where the sum is over $\tau : \{1, 2, \ldots, m\} \to [-m_\epsilon, m_\epsilon]$ compatible with Θ, h, is within ϵ of $\mathcal{A}_L$ and in turn within 2ϵ of $v_{g,h}(\Theta)$.

We now show that $\mathcal{A}_C$ is close to $\mathcal{A}_R$ for a large set of Υ, more precisely a set of μ-measure at least $1 - 2m\epsilon$. For this we need to show:

(i) For a set of Υ of at least μ-measure $1 - \epsilon$ the probability of picking all of the $\sigma(i) \in X_i$ for X_i derived from $\tau : \{1, \ldots, m\} \to [-m_\epsilon, m_\epsilon]$ compatible with Θ, h with $\Upsilon \nvDash \Theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)})$ is $O(\epsilon)$.

_____

[16]We shall shortly use this to impose further properties on r_ϵ.

(ii) For a set of Υ of at least μ-measure $1 - \epsilon$ the probability of picking all of the $\sigma(i) \in X_i$ for X_i derived from $\tau : \{1, \ldots, m\} \to [-m_\epsilon, m_\epsilon]$ *not* compatible with Θ, h with $\Upsilon \models \Theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)})$ is $O(\epsilon)$.

We first tackle (i). By (28), the definitions of K_{0,r_ϵ}, K_{j,r_ϵ} and B_{j,r_ϵ} (pages 15-15), H_j and ϱ (page 12), g and h (pages 16-16) and the definition (2), (3), (4) of compatibility of a colouring τ with Θ, h, for (i) to happen for a particular such τ, it must be the case that one of the following holds:

(a) $\tau(i) = \tau(j) = k \leq 0$, $i \neq j$, $\Theta \models a_i \neq a_j$, $X_i = X_j = K_{k,r_\epsilon}$, $\Upsilon \models a_{\sigma(i)} = a_{\sigma(j)}$.

(b) $\tau(i) = 0$, $\Theta \models F(a_i) = a_i$, $X_i = K_{0,r_\epsilon}$ $\Upsilon \models F(a_{\sigma(i)}) \neq a_{\sigma(i)}$.

Concerning (a), if this were to happen we would have $\sigma(i), \sigma(j) \in H_\zeta$ for some ζ with

$$r_\epsilon^{-1} > {}^\circ(\nu^{-1}|H_\zeta|) > 0.$$

Since r_ϵ has been chosen greater than ϵ^{-1} the probability of making *any* choice of $\sigma(1), \sigma(2), \ldots, \sigma(m)$ such that $\sigma(i), \sigma(j) \in H_\zeta$ for such a H_ζ is less than ϵ. Hence the probability of a type (a) error is less than the number of possible i, j times ϵ.

Concerning (b), this would happen if for some ζ, $\sigma(i) \in H_\zeta$, $\sigma(i) \notin H_{\varrho(\zeta)}$,

$$r_\epsilon^{-1} > {}^\circ(\nu^{-1}|H_\zeta|), \quad r_\epsilon^{-1} > {}^\circ(\nu^{-1}|H_{\varrho(\zeta)}|).$$

From Lemma 13, since r_ϵ is *large*,

$$\mu\{\Upsilon \, : \, |\{j \, : \, F(a_j) = a_j \vee |\{k \, : \, F(a_j) = a_k\}| \geq r_\epsilon^{-1}\nu\}| > \nu(1 - \epsilon)\} > 1 - \epsilon.$$

Hence for μ-measure at least $1 - \epsilon$ the probability of picking such a $\sigma(i)$ is at most ϵ. It follows that for this set of Υ (or Γ) the probability of an error of type (b) is again $O(\epsilon)$ for a set of Υ of measure at least $1 - m\epsilon$.

Now turning to (ii), by the same definitions as for (i), the only ways these errors can occur are if one of the following holds:

(c) $\tau(i) = -k \leq 0$, $i \neq j$, $\Theta \models a_i = a_j$, $X_i = K_{k,r_\epsilon}$, $\Upsilon \models a_{\sigma(i)} = a_{\sigma(j)}$.

(d) $\tau(i) = 0$, $\Theta \models F(a_i) \neq a_i$, $X_i = K_{0,r_\epsilon}$ $\Upsilon \models F(a_{\sigma(i)}) \neq a_{\sigma(i)}$.

By just the same argument as for (i)(a) we can see that the probability of an error of the type (ii)(c) is $O(\epsilon)$. Similarly by the same argument as for (i)(b) for a μ-measure of at most $1 - m\epsilon$ the probability of an error of type (ii)(d) is again $O(\epsilon)$.

Putting all these parts together now we have shown that for $0 < \epsilon \in [0, 1]$ and a set of Υ of μ-measure at least $1 - 2m\epsilon$ the terms in (23) are within a fixed constant times ϵ of each other. The lemma follows.

$\square$

To complete the proof of the Representation Theorem 2 notice that

$$
\begin{aligned}
w(\Theta(a_1,\ldots,a_m)) &= {}^{\circ}w^*(\Theta(a_1,\ldots,a_m)) \\
&= {}^{\circ}\Big(\sum_\Gamma w^*(\Theta \wedge \bigvee \Gamma)\Big) \\
&= {}^{\circ}\Big(\sum_\Gamma w^*(\Theta \,|\, \bigvee \Gamma) \cdot w^*(\bigvee \Gamma)\Big) \\
&= \int_\Gamma {}^{\circ}\big(w^*(\Theta \,|\, \bigvee \Gamma)\big)\, d\mu(\Gamma), \quad \text{by Loeb measure theory,} \\
&\qquad \text{see for example } [5, \text{pages } 17\text{-}20] \text{ or } [6, \text{Chapter } 1] \\
&= \int_{\mathbb{S}_1} v_{g,h}(\Theta)\, d\mu(\Gamma) \quad \text{by Lemma 18.}
\end{aligned}
$$

Conclusions

In our view, including equality and function symbols rather conflicts with the underlying motivation for Pure Inductive Logic. Nevertheless taking the first step in this direction with equality and just one unary function does yield a comprehensible and arguably attractive representation theorem for the probability functions on this language satisfying Ex. A similar result can, we would claim without actually writing out all the details, be shown if we additionally allow finitely many relation and unary function symbols in the language. One can also picture, and conjecture, a similar result even with polyadic function symbols added. However, even if correct, finding a proof in the direction of Theorem 2 currently looks no less than daunting.

We would finally remark that the Representation Theorem 2 we have found has a noticeable similarity to the Representation Theorems given, for example, in [12] for probability functions on polyadic (relational) languages satisfying Spectrum Exchangeability.

References

[1] Carnap, R., The Aim of Inductive Logic, in *Logic, Methodology and Philosophy of Science,* Eds. E.Nagel, P.Suppes & A.Tarski, Stanford University Press, Stanford, California, 1962, pp303-318.

[2] Carnap, R., Inductive Logic and Rational Decisions, in *Studies in inductive logic and probability, Volume I*, Eds. R.Carnap & R.C.Jeffrey, University of California Press, 1971, pp5-31.

[3] Carnap, R., A Basic System of Inductive Logic, in *Studies in Inductive Logic and Probability, Volume I*, Eds. R.Carnap & R.C.Jeffrey, University of California Press, 1971, pp33-165.

[4] Carnap, R., A Basic System of Inductive Logic, in *Studies in Inductive Logic and Probability, Volume II*, Ed. R.C.Jeffrey, University of California Press, 1980, pp7-155.

[5] Cutland, N.J., Loeb Measure Theory, in *Developments in Nonstandard Mathematics*, Eds. N.J.Cutland, F.Oliveira, V.Neves, J.Sousa-Pinto, Pitman Research Notes in Mathematics Series, Vol. 336, Longman Press, 1995, pp151-177.

[6] Cutland, N.J., *Loeb Measures in Practice: Recent Advances,* EMS Lectures, 1997, Springer Lecture Notes in Mathematics, vol.1751.

[7] Gaifman, H., Concerning Measures on First Order Calculi, *Israel Journal of Mathematics*, 1964, **2**:1-18.

[8] Gaifman, H. & Snir, M., Probabilities over Rich Languages, *Journal of Symbolic Logic*, 1982, **47**:495-548.

[9] Howarth, E., *New Rationality Principles in Pure Inductive Logic*, Ph.D. Thesis, Manchester University, June 2015. Available at
http://www.maths.manchester.ac.uk/∼jeff/theses/lwthesis.pdf

[10] Johnson, W.E., Probability: The Deductive and Inductive Problems, *Mind*, 1932, **41**:409-423.

[11] Kuipers, T.A.F., A Survey of Inductive Systems, in *Studies in Inductive Logic and Probability*, (R. C. Jeffrey, editor), vol.II, University of California Press, 1980, pp183-192.

[12] Landes, J., Paris, J.B. & Vencovská, A Survey of some recent results on Spectrum Exchangeability in Polyadic Inductive Logic, *Synthese,* 2011, **181**(1):19-47.

[13] Paris, J.B., *A Short Course in Predicate Logic*, 2014, ISBN 978-87-403-XXXX-X. Available at
http://bookboon.com/en/a-short-course-in-predicate-logic-ebook

[14] Paris, J.B. & Vencovská, A., *Pure Inductive Logic*, in the Association of Symbolic Logic Perspectives in Mathematical Logic Series, Cambridge University Press, April 2015.

[15] Schilpp, P.A. (Ed.), *The Philosophy of Rudolf Carnap,* La Salle, Illinois, Open Court Publishing Company, 1963.