Predicate Exchangeability and Language Invariance in Pure Inductive Logic

Kliess, M.S. and Paris, J.B.

2014

MIMS EPrint: **2015.68**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

ISSN 1749-9097

Predicate Exchangeability and Language Invariance in Pure Inductive Logic*

M.S.Kließ and J.B.Paris
School of Mathematics
The University of Manchester
Manchester M13 9PL

malte.kliess@postgrad.manchester.ac.uk,
jeff.paris@manchester.ac.uk

3rd August 2015

In Pure Inductive Logic, the rational principle of Predicate Exchangeability states that permuting the predicates in a given language L and replacing each occurrence of a predicate in an L-sentence φ according to this permutation should not change our belief in the truth of φ. In this paper we study when a prior probability function w on a purely unary language L satisfying Predicate Exchangeability also satisfies the principle of Unary Language Invariance.

Key words: Predicate Exchangeability, Language Invariance, Inductive Logic, Probability Logic, Uncertain Reasoning.

1 Introduction

In the study of logical probability in the sense of Carnap's Inductive Logic programme, [1], [2], the notion of *symmetry* plays a leading role. In the prior assignment of beliefs, as subjective probabilities, it seems logical, or rational, to observe prevailing symmetries,

*Published in *Logique et Analyse*, **228**:413-540, December 2014.

a typical example being the perceived fairness of a coin toss, at least in the absence of any inside knowledge to the contrary. For this reason a number of rational principles have been proposed in Inductive Logic which are based on invariance under various notions of symmetry, principles which it is argued a choice of logical or rational (we use these two words synonymously) prior probability function should satisfy. I.e. a choice of probability function prior to the acquisition of any evidence, knowledge or intended interpretation. The most prevailing of these principles, accepted by both the founding fathers of Inductive Logic, W.E. Johnson [10], and Rudolf Carnap [3], is that the names we give things, in particular constants and predicates, should not matter when it comes to assigning probabilities. Thus, since interchanging which side of the coin we call heads and which we call tails does not change what we understand by a coin toss, both outcomes should rationally receive the same probability.

A second, ubiquitous, rational principle is that when assigning rational probabilities 'irrelevant information' can be disregarded. Indeed the central principle of Johnson and Carnap, the so called *Johnson's Sufficientness Postulate*, is just such an example. Just as with saying what exactly we might mean by a 'symmetry' this directive does of course raise the question of what exactly we mean by an 'irrelevant information', and numerous interpretations have been mooted, generally based on the idea that such information is expressed in a disjoint, or partially disjoint language.

A third, more recent and rather overarching, rational principle is the requirement of *language invariance*. By that we mean that to be rational a probability function should not be restricted to one special language but be extendable to larger languages, and furthermore that those additional rational principles which we imposed in the context of the original language should also be satisfied by these extensions.

In this paper we shall study two symmetry principles, *Constant Exchangeability*[1] and *Predicate Exchangeability*, in the presence of language invariance with the main goal of providing a representation theorem along the lines of de Finetti's Representation Theorem for Constant Exchangeability alone, see for example [5], [11]. Although rather technical, at least in relation to the seemingly elementary mathematics at the heart of Inductive Logic, such results have, starting with Gaifman [6] and Humburg [9], been an extremely powerful tool in our understanding of the interrelationship between the various rational principles which have been proposed. Hopefully the results given here will also find similar applications in the future.

The structure of this paper is as follows. In Section 2 we shall introduce the notation and give precise formulations of the main principles we shall be studying. In Section 3 we shall provide a representation theorem for probability functions satisfying language invariance with Constant and Predicate Exchangeability assuming a particularly strong irrelevance condition, the *Constant Irrelevance Principle*, and in the next section show a similar result without this assumption. This latter representation theorem shows that all such

[1]Johnson's *Permutation Postulate* and Carnap's *Axiom of Symmetry.*

probability functions are in a sense convex mixtures of probability functions satisfying the so called *Weak Irrelevance Principle*, and conversely. Finally in Section 5 we will give a general representation theorem for probability functions satisfying Constant and Predicate Exchangeability alone, showing that they are mixtures (not necessarily convex) of such probability functions which additionally satisfy language invariance.

The philosophical standpoint of this paper is *Pure Inductive Logic*, see [11], [12], a branch of Carnap's Inductive Logic which he already described in [3]. Thus we shall be interested in studying the (prior) assignment of logical probability in the absence of any acquired evidence or knowledge and without relation to any specific interpretations. Of course the rational principles that one might consider imposing on this assignment may have their genesis in real world examples but once a principle is formulated it is studied in Pure Inductive Logic through the agency of mathematics. The subsequent value to philosophy lies, we would opine, mainly in the philosophically interesting conclusions that this mathematical investigation engenders. In other words the aim is to explicate the philosophical consequences of making certain philosophically motivated assumptions via the method of rigorous mathematical proof. The fact that the necessary mathematics linking two philosophically approachable assertions may be rather technical is clearly not ideal but nevertheless that currently seems on occasions to be unavoidable.

2 Notation and Principles

We will be working in the usual context of (unary) Pure Inductive Logic. Thus the first order languages we will be concerned with consist only of finitely many unary predicate symbols P_i and countably many constant symbols[2] $a_1, a_2, \ldots, a_m, \ldots$, which should be thought of as exhausting the universe. We will write L_q to indicate the language containing just the predicates $P_1, \ldots, P_q$. Let SL denote the set of sentences of the language L, $QFSL$ the set of quantifier-free sentences of L.

An *atom* $\alpha(x)$ of L is a formula

$$P_1^{\varepsilon_1}(x) \wedge P_2^{\varepsilon_2}(x) \wedge \cdots \wedge P_q^{\varepsilon_q}(x),$$

with $\varepsilon_i \in \{0,1\}$ and $P_i^1(x)$, $P_i^0(x)$ standing for $P_i(x)$, $\neg P_i(x)$, respectively.[3] Note that for L containing q predicates there are 2^q atoms, which we shall denote $\alpha_1, \ldots, \alpha_{2^q}$.

A *state description* of L for[4] $a_{i_1}, \ldots, a_{i_n}$ is a sentence

$$\Theta(a_{i_1}, \ldots, a_{i_n}) = \bigwedge_{j=1}^{n} \alpha_{h_j}(a_{i_j}),$$

[2]For convenience, we shall henceforth refer to these just as 'predicates' and 'constants'.

[3]In the literature, the notation $\pm P_i(x)$ is more common; however, in the scope of this paper, the notation $P_i^{\varepsilon_i}(x)$ is more convenient.

[4]The entries in such lists will be taken to be distinct unless otherwise stated.

where $h_j \in \{1, \ldots, 2^q\}$ for $j = 1, \ldots, n$.

A *probability function on* L is a function $w : SL \to [0,1]$ satisfying the following conditions for all $\vartheta, \varphi, \exists x\, \psi(x) \in SL$:

(P1) If $\models \vartheta$, then $w(\vartheta) = 1$.

(P2) If $\vartheta \models \neg\varphi$, then $w(\vartheta \vee \varphi) = w(\vartheta) + w(\varphi)$.

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

The following theorem will allow us to restrict our studies to quantifier-free sentences.

Theorem 1 (Gaifman, [7]). *Let $w : QFSL \to [0,1]$ be a function satisfying (P1), (P2) for all $\vartheta, \varphi \in QFSL$. Then there exists a unique $w' : SL \to [0,1]$ satisfying (P1)-(P3) extending w.*

Since any quantifier-free sentence of L is logically equivalent to a disjunction of state descriptions, by (P2) and Theorem 1 a probability function is determined by its values on the state descriptions. Let $\vec{x} \in \mathbb{D}_{2^q}$, where

$$\mathbb{D}_{2^q} := \left\{ \langle x_1, \ldots, x_{2^q} \rangle \mid x_i \geq 0, \sum_{i=1}^{2^q} x_i = 1 \right\}.$$

Noticing that atoms instantiated by different constants are logically independent we can obtain an example of a probability function $w_{\vec{x}}$ by treating them as even stochastically independent and defining $w_{\vec{x}}$ on state descriptions via

$$w_{\vec{x}}(\Theta(a_{i_1}, \ldots, a_{i_n})) := \prod_{i=1}^{2^q} x_i^{n_i},$$

where $n_i = |\{j \mid h_j = i\}|$.

These functions are quite important examples, as they form the building blocks in de Finetti's Representation Theorem. Before stating this theorem, we need to introduce the Principle of Constant Exchangeability:

The Principle of Constant Exchangeability, Ex
A probability function w on SL satisfies Constant Exchangeability *if for each $\varphi(a_1, \ldots, a_n) \in SL$, and σ a permutation of* $\mathbb{N}^+$ $(= \{1, 2, 3, \ldots\})$,

$$w(\varphi(a_1, \ldots, a_n)) = w(\varphi(a_{\sigma(1)}, \ldots, a_{\sigma(n)})).$$

The justification for Ex as a principle of rationality is based on a symmetry argument. That there is complete symmetry between the constants and hence that to ascribe different probabilities to $w(\varphi(a_1, \ldots, a_n))$ and $w(\varphi(a_{\sigma(1)}, \ldots, a_{\sigma(n)}))$ would therefore be irrational.

Notice that the $w_{\vec{x}}$ satisfy Ex. Ex is such a well accepted principle in Inductive Logic that we shall henceforth take it as a standing assumption throughout that all the probability functions we consider satisfy it.

We shall therefore not mention the particular constants whenever they are understood from the context.

Theorem 2 (de Finetti's Representation Theorem). *Let $L = L_q$ and w be a probability function on SL satisfying Ex. Then there exists a normalized, σ-additive measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ such that*

$$w\left(\bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\left(\bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) d\mu(\vec{x}). \tag{1}$$

Conversely, given such a measure μ, the function w defined by (1) *is a probability function on SL satisfying Ex.*

It is straightforward to show (see [12]) that these $w_{\vec{x}}$ are characterized as those probability functions which satisfy Ex together with

The Principle of Constant Irrelevance, IP
A probability function w on SL satisfies Constant Irrelevance *if for $\vartheta, \varphi \in QFSL$ with no constants in common,*

$$w(\vartheta \wedge \varphi) = w(\vartheta) \cdot w(\varphi).$$

Thus de Finetti's Representation Theorem can be alternately stated as saying that every probability function satisfying Ex is a convex mixture of probability functions satisfying IP, and conversely.

The principles that are of particular interest to us in this paper are:

The Principle of Predicate Exchangeability, Px
A probability function w on SL satisfies Predicate Exchangeability *if whenever $\varphi \in SL$ and φ' is the result of replacing the predicates*[5] *$P_{i_1}, \dots, P_{i_m}$ in φ by $P_{k_1}, \dots, P_{k_m}$, then*

$$w(\varphi) = w(\varphi').$$

The justification for this principle is just the same as for Constant Exchangeability, as in presence of no prior knowledge of the universe a rational agent should not favour any particular predicate in her language from the start.

[5] In such lists we shall always assume that the members are distinct.

As a motivational example[6] suppose a rational agent is picking (with replacement) balls from an urn, these balls being either black or white and either shiny or matt, and she is asked about her subjective probabilities concerning the distribution of colours and textures of balls in this urn. Interpreting the constants of the language as the balls being picked from the urn and two predicate of the language as describing a ball's colour and texture, it would seem irrational that in the absence of any prior knowledge of the urn the agent should assign a higher probability to 'white' than to 'shiny', say.

The Principle of Unary Language Invariance, ULi
A probability function w on SL satisfies Unary Language Invariance *if there exists a family of probability functions $w^{\mathcal{L}}$, one for each finite (unary) language $\mathcal{L}$, satisfying Px (and by standing assumption Ex), such that $w = w^L$ and whenever $\mathcal{L}' \subseteq \mathcal{L}$, then $w^{\mathcal{L}'} = w^{\mathcal{L}} \restriction S\mathcal{L}'$, the restriction of w to the sentences of $\mathcal{L}'$.*

We say that w satisfies ULi with $\mathcal{P}$ *(for some principle $\mathcal{P}$), if each of the functions $w^{\mathcal{L}}$ satisfy $\mathcal{P}$.*

This principle allows a rational agent to extend her language should the situation require it. For instance, suppose that in the aforementioned urn example the agent were to find that after picking a number of balls and noting their colours and textures she was to discover that some of the balls were magnetized and others not. She might not already have a predicate interpreted as 'is magnetized' in her language, and so might want to add one in order to further distinguish the balls. However, upon learning that some balls were magnetized it would surely be irrational to discard all the properties noted about the previously selected balls. If the agent's probability function satisfies ULi she will be able to just extend the language to include an additional 'is magnetized' predicate, which by Px would initially have had precisely the same status as the colour and texture predicates, and consequently she could just continue without having to start over again. After all, just because some balls are magnetized does not mean the agent has reason to change her belief about the colour and texture distribution of the balls in the urn.

The 'rationality' of ULi is based on two considerations. Firstly the idea that if the agent chooses probability functions $w^{\mathcal{L}}$ and $w^{\mathcal{L}'}$ on languages $\mathcal{L}, \mathcal{L}'$ respectively and $\mathcal{L} \subset \mathcal{L}'$ then $w^{\mathcal{L}'}$ restricted to $S\mathcal{L}$ should agree with $w^{\mathcal{L}}$, otherwise she would be in the seemingly irrational position of giving different probabilities to the same sentence simply on account of their being other unmentioned predicate symbols in the language. Put another way her choice $w^{\mathcal{L}}$ would depend on the particular set of predicates in $\mathcal{L}$, and she would be imposing some a priori semantics on the languages.

The second consideration is that if our agent subscribes to some principles $\mathcal{P}$ as rational obligations then this should not be a function of the particular language under consid-

[6]We need to be constantly on our guard with such examples. It is crucially important to appreciate that they are intended only to motivate an underlying arguably rational principle, in this case Predicate Exchangeability, not to propose a practical rule applicable to all interpretations of the language. The widespread failure to appreciate this point has proved the bane of this subject.

eration, so that the agent should be subscribing to $\mathcal{P}$ for any language. Taking $\mathcal{P}$ to be Px (and the standing Ex) these two considerations give ULi.[7,8]

Notice that if $w^{\mathcal{L}}, w^{\mathcal{L}'}$ are members of a language invariant family and $\mathcal{L}, \mathcal{L}'$ have the same number of predicates then $w^{\mathcal{L}}$ is the same as $w^{\mathcal{L}'}$ up to renaming predicates. For that reason it will, for the most part, be enough for us to focus our attention on the members $w^{\mathcal{L}}$ of the family when $\mathcal{L} = L_q$ for some q.

Given a permutation σ of the predicates of L, there is a unique permutation of the atoms of L that is induced by σ: For $\alpha(x) = \bigwedge_{i=1}^{q} P_i^{\varepsilon_i}(x)$ an atom of L, let $\sigma\alpha(x)$ be the atom given by

$$\sigma\alpha(x) = \bigwedge_{i=1}^{q} \sigma(P_i)^{\varepsilon_i}(x).$$

This now in turn induces a permutation on SL in the obvious way. Abusing notation, we identify these permutations of atoms and L-sentences with σ. We shall write *σ is induced by Px* to indicate that σ arises from a permutation of predicates.

3 A First Representation Theorem

Since the $w_{\vec{x}}$ are the building blocks for probability functions satisfying Ex (see de Finetti's Theorem above), these functions are of special interest to us. We will therefore begin by studying when they satisfy ULi, equivalently when probability functions satisfying Ex and IP satisfy ULi.

Suppose a probability function w on some language L satisfies Predicate Exchangeability. Then the probability that w assigns any atom α of L only depends on the number of predicates in α that occur negated.[9] To see this notice that if α, α' are atoms then α' can be obtained from α by a permutation of predicates just if both atoms have the same number of negated predicates.

It is thus convenient to introduce a function assigning each atom the corresponding number of predicates:

Definition 3:
Let $L = L_q$. Define $\gamma_q : \{1, \ldots, 2^q\} \to \{0, \ldots, q\}$ by

$$\gamma_q(i) = k \Leftrightarrow \alpha_i \text{ contains } k \text{ negated predicates.}$$

[7] In fact, as one easily checks, without Px, any probability function w can be arbitrarily extended to obtain a language invariant family, which makes Language Invariance in this form a trivial statement.

[8] It is interesting to note, as pointed out by one of the referees, that language invariance is a wholly unobjectionable feature of logical consequence so it's continued acceptance for our more general form of reasoning would seem in the first instance entirely natural.

[9] This is an arbitrary choice. One could also count the number of predicates that occur positively in α, as the argument is symmetrical.

We shall drop the index q whenever it is understood from the context.

Now considering $\vec{c} \in \mathbb{D}_{2^q}$ it follows that $w_{\vec{c}}$ satisfies Predicate Exchangeability if and only if $c_i = c_j$ whenever $\gamma(i) = \gamma(j)$. With this in mind we shall assume that our enumeration of the atoms is such that the number of negated predicates is non-decreasing as we move right through $\alpha_1, \alpha_2, \ldots, \alpha_{2^q}$. Since for each $i \in \{0, \ldots, q\}$ there are $\binom{q}{i}$ atoms of L_q with i predicates occurring negatively we therefore have that for $w_{\vec{c}}$ satisfying Px

$$\vec{c} = \langle \mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_2, \ldots, \mathcal{C}_{q-1}, \ldots, \mathcal{C}_{q-1}, \mathcal{C}_q \rangle,$$

i.e. $c_i = \mathcal{C}_{\gamma(i)}$ for $i = 1, 2, \ldots, 2^q$, and

$$\sum_{i=0}^{q} \binom{q}{i} \mathcal{C}_i = 1.$$

Thus any such $\vec{c}$ gives us a unique $\vec{\mathcal{C}} = \langle \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_q \rangle$ with the properties

$$\forall i \in \{0, \ldots, q\}\ \mathcal{C}_i \geq 0 \text{ and } 1 = \sum_{i=0}^{q} \binom{q}{i} \mathcal{C}_i.$$

Conversely, any $\vec{\mathcal{C}}$ with these properties provides a unique $\vec{c} \in \mathbb{D}_{2^q}$ such that $w_{\vec{c}}$ satisfies Px, giving us a 1-1 correspondence between these $\vec{c} \in \mathbb{D}_{2^q}$ and the elements of

$$\widehat{\mathbb{D}}_q := \left\{ \vec{\mathcal{C}} = \langle \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_q \rangle \mid \forall i \in \{0, \ldots, q\}\, \mathcal{C}_i \geq 0 \text{ and } 1 = \sum_{i=0}^{q} \binom{q}{i} \mathcal{C}_i \right\}. \quad (2)$$

We shall refer to elements of the set above as the *alternative notation* for such a $\vec{c} \in \mathbb{D}_{2^q}$.

Given an atom α of L_q, we can view this atom as a quantifier-free sentence in the extended language L_{q+1}, and obtain

$$\alpha(x) \equiv \alpha^+(x) \vee \alpha^-(x) = (\alpha(x) \wedge P_{q+1}(x)) \vee (\alpha(x) \wedge \neg P_{q+1}(x)).$$

Now suppose $\vec{c} \in \mathbb{D}_{2^q}$, $\vec{d} \in \mathbb{D}_{2^{q+1}}$ are such that $w_{\vec{d}} \restriction SL_q = w_{\vec{c}}$ and both satisfy Px. Then by the logical equivalence given above, we must have

$$w_{\vec{c}}(\alpha) = w_{\vec{d}}(\alpha) = w_{\vec{d}}(\alpha^+) + w_{\vec{d}}(\alpha^-).$$

Suppose $\vec{\mathcal{C}} \in \widehat{\mathbb{D}}_q$, $\vec{\mathcal{D}} \in \widehat{\mathbb{D}}_{q+1}$ are the corresponding alternative notations for $\vec{c}$ and $\vec{d}$. Then we obtain for each $i \in \{0, \ldots, q\}$,

$$\mathcal{C}_i = \mathcal{D}_i + \mathcal{D}_{i+1}.$$

The following proposition generalizes this to ULi families.

Proposition 4. *Let $w_{\vec{c}}$ be a probability function on L_q. Suppose $w_{\vec{c}}$ is a member of a ULi with IP family $\mathcal{W}$ and assume $w_{\vec{d}} \in \mathcal{W}$ is a probability function on L_r for some $r > q$. Let $\vec{\mathcal{C}}, \vec{\mathcal{D}}$ be the corresponding alternative notations for $\vec{c}$, $\vec{d}$. Then for each $j \in \{0, \dots, q\}$, we have*

$$\mathcal{C}_j = \sum_{k=j}^{r-q+j} \binom{r-q}{k-j} \mathcal{D}_k.$$

Proof: We show this by induction on $s := r - q$. In case $s = 1$, we have for each $j \in \{0, \dots, q\}$,

$$\mathcal{C}_j = \mathcal{D}_j + \mathcal{D}_{j+1},$$

since for α an atom of L_q with j negated predicates, we have in L_r $(= L_{q+1})$

$$\alpha = \alpha^+ \vee \alpha^-,$$

where α^+, α^- are atoms of L_r with $j, j+1$ negated predicates, respectively.

Now let $s = p + 1$ and assume the result holds for p. Let $\mathcal{D}'_i$ denote the corresponding values for the atoms of L_{q+p}. By the inductive hypothesis we have

$$\mathcal{C}_j = \sum_{k=j}^{(q+p)-q+j} \binom{(q+p)-q}{k-j} \mathcal{D}'_k.$$

Just as in the case $s = 1$ we have $\mathcal{D}'_k = \mathcal{D}_k + \mathcal{D}_{k+1}$ for each $0 \le k \le q + p$, so we obtain

$$\mathcal{C}_j = \sum_{k=j}^{p+j} \binom{p}{k-j} (\mathcal{D}_k + \mathcal{D}_{k+1}) = \sum_{k=j}^{p+1+j} \binom{p+1}{k-j} \mathcal{D}_k = \sum_{k=j}^{r-q+j} \binom{r-q}{k-j} \mathcal{D}_k,$$

as required. ⊣

With this proposition in mind, we are ready to proceed to our first Representation Theorem.

Theorem 5. *Let $\vec{c} \in \mathbb{D}_{2^q}$ and $w_{\vec{c}}$ be a probability function satisfying Px. Then $w_{\vec{c}}$ is a member of a ULi with IP family $\mathcal{W} = \{w_{\vec{d_r}} \mid \vec{d_r} \in \mathbb{D}_{2^r}\}$ if and only if each entry c_i of $\vec{c}$ is of the form*

$$c_i = \int_{[0,1]} x^{\gamma(i)} (1-x)^{q-\gamma(i)} \, d\rho(x) \tag{3}$$

for some normalized σ-additive measure ρ on $[0, 1]$.

Proof: We will use methods from Nonstandard Analysis working in a suitable nonstandard universe ${}^*V$, see for example [4]. The key idea to the proof is to marginalize some $w_{\vec{c}}$ on some infinite language to finite languages, rather than constructing extensions of some $w_{\vec{d}}$ on a finite language to each finite level. Suppose we have such a ULi with IP family $\mathcal{W}$ of probability functions, so for each $r \in \mathbb{N}$, we have some $w^{(r)}$ on L_r in this family. By the Transfer Principle this holds for each $r \in {}^*\mathbb{N}$, so we can pick some nonstandard natural number $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$ and consider $w^{(\nu)}$. Now $w^{(\nu)} \restriction SL_r = w^{(r)}$ for each $r < \nu$, as these are members of the same ULi family and we can retrieve our original family $\mathcal{W}$ by looking at functions of the form $w^{(\nu)} \restriction SL_r$ for $r \in \mathbb{N}$, taking standard parts – denoted as usual by $^\circ$ – where necessary.

In more detail let ${}^*V$ be a nonstandard universe that contains at least $\mathbb{D}_{2^q}$ for finite $q \in \mathbb{N}$, all probability functions $w_{\vec{b}}$ satisfying Px and everything else needed in this proof. Let $\nu \in {}^*\mathbb{N}$ be nonstandard and consider $\vec{b} \in \mathbb{D}_{2^\nu}$ such that $w_{\vec{b}}$ on L_ν satisfies Px. Assume that $\vec{\mathcal{B}}$ is the alternative notation for $\vec{b}$ given by (2). For each $q < \nu$, we can define a probability function on L_q *in* ${}^*V$ satisfying Px by letting

$$\mathcal{C}_j = \sum_{\kappa=j}^{\nu-q+j} \binom{\nu-q}{\kappa-j} \mathcal{B}_\kappa$$

for $j = 0, \ldots, q$. In general, this gives $\vec{c} \in {}^*\mathbb{D}_{2^q}$, so we need to take the standard part of $\vec{c}$, denoted $^\circ\vec{c}$, to get a probability function $w_{^\circ\vec{c}}$ in V.

We will first look at $\vec{\mathcal{B}}$ when all weight is concentrated on a single $\mathcal{B}_\kappa$, $0 \leq \kappa \leq \nu$. Since we need to have $\sum_{\kappa=0}^{\nu} \binom{\nu}{\kappa} \mathcal{B}_\kappa = 1$, we obtain

$$\mathcal{B}_\kappa = \binom{\nu}{\kappa}^{-1}.$$

Then we get for $0 \leq j \leq q$

$$\begin{aligned}
\mathcal{C}_j = \binom{\nu-q}{\kappa-j} \mathcal{B}_\kappa &= \binom{\nu-q}{\kappa-j} \cdot \binom{\nu}{\kappa}^{-1} \\
&= \frac{(\nu-q)! \cdot \kappa! \cdot (\nu-\kappa)!}{(\kappa-j)! \cdot (\nu-q-\kappa+j)! \cdot \nu!} \\
&= \frac{\kappa \cdot (\kappa-1) \cdots (\kappa-j+1) \cdot (\nu-\kappa) \cdots (\nu-\kappa-q+j+1)}{\nu \cdot (\nu-1) \cdots (\nu-q+1)},
\end{aligned} \tag{4}$$

thus leading to the standard part being

$$^\circ\mathcal{C}_j = {}^\circ\left(\left(\frac{\kappa}{\nu}\right)^j \cdot \left(1-\frac{\kappa}{\nu}\right)^{q-j}\right) = {}^\circ\left(\frac{\kappa}{\nu}\right)^j \cdot \left(1 - {}^\circ\left(\frac{\kappa}{\nu}\right)\right)^{q-j}. \tag{5}$$

Now consider an arbitrary $\vec{\mathcal{B}} = \langle \mathcal{B}_0, \ldots, \mathcal{B}_\nu \rangle$. Then for each $0 \leq \kappa \leq \nu$ there exists

$\gamma_\kappa \in {}^*[0,1]$ such that we can write

$$\mathcal{B}_\kappa = \gamma_\kappa \cdot \binom{\nu}{\kappa}^{-1}.$$

Note that since

$$\sum_{\kappa=0}^{\nu} \binom{\nu}{\kappa} \mathcal{B}_\kappa = 1$$

we must have

$$\sum_{\kappa=0}^{\nu} \gamma_\kappa = 1.$$

Then using (4) we see that each summand in $\mathcal{C}_j$ will be of the form

$$\gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1},$$

thus ${}^\circ\mathcal{C}_j$ will become

$${}^\circ\mathcal{C}_j = {}^\circ\left(\sum_{\kappa=j}^{\nu-q+j} \gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right).$$

Since we are only interested in the standard part, we can add the finitely many summands for $\kappa = 0, \ldots, j-1, \nu-q+j+1, \ldots, \nu$ without changing ${}^\circ\mathcal{C}_j$ (assuming that $0 < j < q$), as we have

$$\begin{aligned}
&{}^\circ\left(\sum_{\kappa=0}^{\nu} \gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right) - {}^\circ\mathcal{C}_j \\
&= {}^\circ\left(\sum_{\kappa=0}^{j-1} \gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right) + {}^\circ\left(\sum_{\kappa=\nu-q+j+1}^{\nu} \gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right) \\
&= \sum_{\kappa=0}^{j-1} {}^\circ\left(\gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right) + {}^\circ\left(\sum_{\kappa=\nu-q+j+1}^{\nu} \gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right) \\
&= 0 + 0,
\end{aligned}$$

because for $\kappa \in \{0, \ldots, j-1, \nu-q+j+1, \ldots, \nu\}$, either ${}^\circ(\kappa/\nu) = 0$ or ${}^\circ(1-\kappa/\nu) = 0$, so the first and last sum vanish as each consists of finitely many terms. Note that in case $j = 0, q$, either the first or the second summand is empty, and therefore we can apply the same argument for $j = 0, q$ as well, giving

$${}^\circ\mathcal{C}_j = {}^\circ\left(\sum_{\kappa=0}^{\nu} \gamma_\kappa \cdot \binom{\nu-q}{\kappa-j}\binom{\nu}{\kappa}^{-1}\right) \tag{6}$$

for $j \in \{0, \ldots, q\}$.

Now let $N = \{0, \ldots, \nu\}$ and (in ${}^*V$ of course) let μ be the Loeb counting measure on N (see example (1), section 2 in [4]). Then we can write (6) as

$$^\circ\mathcal{C}_j = {}^\circ\!\!\int_N \gamma_\kappa \cdot \binom{\nu - q}{\kappa - j}\binom{\nu}{\kappa}^{-1} d\mu(\kappa). \tag{7}$$

Let μ' be the discrete measure on ${}^*[0,1]$ which for $\kappa \in N$ gives the point κ/ν measure γ_κ. Then we get

$$\int_N \gamma_\kappa \cdot \binom{\nu - q}{\kappa - j} \cdot \binom{\nu}{\kappa}^{-1} d\mu(\kappa) = \int_{{}^*[0,1]} \binom{\nu - q}{x \cdot \nu - j} \cdot \binom{\nu}{x \cdot \nu}^{-1} d\mu'(x). \tag{8}$$

Now let ρ be the measure *in* V on $[0,1]$ which for a Borel subset A of $[0,1]$ gives

$$\rho(A) = {}^\circ\mu'({}^*A). \tag{9}$$

By well known results from Loeb Measure Theory, see for example [4],

$${}^\circ\!\!\int_{{}^*[0,1]} \binom{\nu - q}{x \cdot \nu - j} \cdot \binom{\nu}{x \cdot \nu}^{-1} d\mu'(x) = \int_{[0,1]} {}^\circ\!\left(\binom{\nu - q}{x \cdot \nu - j} \cdot \binom{\nu}{x \cdot \nu}^{-1}\right) d\rho(x). \tag{10}$$

Combining (5),(7),(8),(10) now gives that

$$^\circ\mathcal{C}_j = \int_{[0,1]} x^j \cdot (1 - x)^{q-j}\, d\rho(x) \tag{11}$$

We obtain a $\vec{c} \in \mathbb{D}_{2^q}$ by letting

$$\vec{c} = \langle {}^\circ\mathcal{C}_0, {}^\circ\mathcal{C}_1, \ldots, {}^\circ\mathcal{C}_1, \ldots, {}^\circ\mathcal{C}_{q-1}, \ldots, {}^\circ\mathcal{C}_{q-1}, {}^\circ\mathcal{C}_q \rangle.$$

As we can marginalize $\vec{b}$ in the above way to any $r \in \mathbb{N}$, we obtain that given a family of functions $\{w_{\vec{d_r}} \mid \vec{d_r} \in \mathbb{D}_{2^r}\}$ such that each $\vec{d_r}$ is obtained by marginalizing some $\vec{b} \in \mathbb{D}_{2^\nu}$ and therefore satisfies (3), this family satisfies Unary Language Invariance.

For the converse it is straightforward to check that any $w_{\vec{c}}$ for which all the c_i in $\vec{c}$ are of the form (11) does satisfy ULi, the required family member on L_r being obtained simply by changing q to r with the same measure ρ. $\dashv$

However, as the following example will show, the probability functions of the form $w_{\vec{c}}$ satisfying ULi with IP are not the building blocks that generate all probability functions satisfying ULi:

Example 6. Let $c_0^{L_2}$ be the probability function on L_2 given by

$$c_0^{L_2} = \frac{1}{4}\left(w_{\langle 1,0,0,0\rangle} + w_{\langle 0,1,0,0\rangle} + w_{\langle 0,0,1,0\rangle} + w_{\langle 0,0,0,1\rangle}\right).$$

Then $c_0^{L_2}$ satisfies ULi as it is a member of Carnap's Continuum of Inductive Methods (see e.g. [12]). However, both $\langle 0,1,0,0\rangle$ and $\langle 0,0,1,0\rangle$ are not of the form (3), and thus $c_0^{L_2}$ shows that we cannot have a Representation Theorem for w satisfying ULi of the form

$$w = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\, d\mu(\vec{x})$$

with μ giving all weight to $\vec{c}$ of the form (3).

4 The Representation Theorem for w satisfying ULi

In the previous section, we used a probability function satisfying Px + IP on the infinite language L_ν to construct a language invariant family by marginalizing to each finite level. In turn this gave us our first Representation Theorem, Theorem 5.

In this section we shall instead derive a representation theorem for just ULi by using an arbitrary state description Υ of L_ν to construct a probability function satisfying Px by averaging over all permutations of predicates, similarly to the definition of $c_0^{L_2}$ in Example 6. We first introduce some notation and a related result, Theorem 8, which is of interest in its own right.

Let $\Upsilon(P_1,\dots,P_\nu,a_1,\dots,a_\nu)$ be the state description of L_ν given by

$$\Upsilon(P_1,\dots,P_\nu,a_1,\dots,a_\nu) = \bigwedge_{i=1}^{\nu}\bigwedge_{j=1}^{\nu} P_i^{\varepsilon_{i,j}}(a_j).$$

Then we can represent Υ by the $\nu\times\nu$ - matrix

$$\begin{pmatrix} \varepsilon_{1,1} & \varepsilon_{1,2} & \cdots & \varepsilon_{1,\nu} \\ \varepsilon_{2,1} & \varepsilon_{2,2} & \cdots & \varepsilon_{2,\nu} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{\nu,1} & \varepsilon_{\nu,2} & \cdots & \varepsilon_{\nu,\nu} \end{pmatrix}.$$

Now consider the $q\times\nu$ - matrix Ψ where the j'th row of Ψ is the i_j'th row of Υ, for some $i_1,\dots,i_q \in \{1,\dots,\nu\}$, not necessarily distinct. Then we can similarly think of Ψ as a state description $\Psi(a_1,\dots,a_\nu)$ of L_q. So each column of Ψ represents an atom of L_q, and we obtain $\vec{c}\in {}^*\mathbb{D}_{2^q}$ by letting

$$c_i = \frac{|\{j \mid \Psi \models \alpha_i(a_j)\}|}{\nu}.$$

We thus obtain for each $\langle i_1, \ldots, i_q \rangle$ with $1 \le i_1, \ldots, i_q \le \nu$ some $w_{\vec{c}}$ for $\vec{c} \in {}^*\mathbb{D}_{2^q}$, which we shall denote by $w^{\Upsilon}_{\langle i_1, \ldots, i_q \rangle}$.

We can now define the functions that we will then use to prove the representation theorem for general ULi functions.

Definition 7:
Let $\Upsilon(P_1, \ldots, P_\nu, a_1, \ldots, a_\nu)$ be a state description of L_ν for ν distinct constants. Let $L = L_q$ for some finite q. For $i_1, \ldots, i_q \in \{1, \ldots, \nu\}$, not necessarily distinct, let $w^{\Upsilon}_{\langle i_1, \ldots, i_q \rangle}$ be given as above.

Define the function ∇^L_Υ on SL by

$$\nabla^L_\Upsilon = \sum_{e:\{1,\ldots,q\}\to\{1,\ldots,\nu\}} \frac{1}{\nu^q} w^{\Upsilon}_{\langle e(1), \ldots, e(q) \rangle}.$$

Instead of just marginalizing to the first q rows, as we did in the case of $w_{\vec{c}}$, ∇^L_Υ now also averages over all permutations of the predicates. One can think of this as picking q rows from the matrix representing Υ *with replacement* to obtain the predicates $P_1, \ldots, P_q$ of L_q.

Before our next result we need to recall another principle, see [8], [12].

The Weak Irrelevance Principle, WIP
A probability function w on SL satisfies Weak Irrelevance *if whenever $\vartheta, \varphi \in QFSL$ have no constants nor predicates in common then*

$$w(\vartheta \wedge \varphi) = w(\vartheta) \cdot w(\varphi).$$

Notice that this is a weakening of the Constant Irrelevance principle, IP, where we required only that ϑ, φ have no constants in common.

Theorem 8. *Let $\Upsilon(P_1, \ldots, P_\nu, a_1, \ldots, a_\nu)$ be a state description of L_ν and let $L = L_q$. Then the function ${}^\circ\nabla^L_\Upsilon$ is (can be extended to) a probability function on SL satisfying ULi + WIP.*

Proof: From the definition of ∇^L_Υ it is obvious that ${}^\circ\nabla^L_\Upsilon$ is a probability function satisfying Ex.

For Px, let σ be a permutation of the predicates of L. Then we obtain

$$\begin{aligned} {}^\circ\nabla^L_\Upsilon(\sigma\Theta) &= {}^\circ\left[\sum_{e:\{1,\ldots,q\}\to\Upsilon} \frac{1}{\nu^q} \cdot w^{\Upsilon}_{\langle e(1), \ldots, e(q) \rangle}(\sigma\Theta)\right] \\ &= {}^\circ\left[\sum_{e:\{1,\ldots,q\}\to\Upsilon} \frac{1}{\nu^q} \cdot w^{\Upsilon}_{\langle e(\sigma^{-1}(1)), \ldots, e(\sigma^{-1}(q)) \rangle}(\Theta)\right], \end{aligned}$$

since σ permutes the predicates of L,

$$
\begin{aligned}
&= {}^{\circ}\left[\sum_{e\circ\sigma^{-1}:\{1,\ldots,q\}\to\Upsilon} \frac{1}{\nu^q}\cdot w^{\Upsilon}_{\langle e\circ\sigma^{-1}(1),\ldots,e\circ\sigma^{-1}(q)\rangle}(\Theta)\right] \\
&= {}^{\circ}\left[\sum_{e':\{1,\ldots,q\}\to\Upsilon} \frac{1}{\nu^q}\cdot w^{\Upsilon}_{\langle e'(1),\ldots,e'(q)\rangle}(\Theta)\right] = {}^{\circ}\nabla^{L}_{\Upsilon}(\Theta).
\end{aligned}
$$

To show that ULi holds, notice that for $\Theta(a_1,\ldots,a_n)$ the state description

$$\Theta(a_1,\ldots,a_n) = \bigwedge_{j=1}^{n} \alpha_{h_j}(a_j),$$

we obtain on L_{q+1},

$$\Theta(a_1,\ldots,a_n) = \bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha^{\varepsilon_j}_{h_j}(a_j),$$

where

$$\alpha^{\varepsilon_j}_{h_j}(x) = \alpha_{h_j}(x)\wedge P^{\varepsilon_j}_{q+1}(x).$$

We obtain

$$
\begin{aligned}
&{}^{\circ}\nabla^{L_{q+1}}_{\Upsilon}(\Theta) \\
&\quad = \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} {}^{\circ}\nabla^{L_{q+1}}_{\Upsilon}\left(\bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha^{\varepsilon_j}_{h_j}\right) \\
&\quad = \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} {}^{\circ}\left[\sum_{e:\{1,\ldots,q+1\}\to\{1,\ldots,\nu\}} \frac{1}{\nu^{q+1}} w^{\Upsilon}_{\langle e(1),\ldots,e(q+1)\rangle}\left(\bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha^{\varepsilon_j}_{h_j}\right)\right] \\
&\quad = {}^{\circ}\left[\sum_{e:\{1,\ldots,q+1\}\to\{1,\ldots,\nu\}} \frac{1}{\nu^{q+1}} \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} w^{\Upsilon}_{\langle e(1),\ldots,e(q+1)\rangle}\left(\bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha^{\varepsilon_j}_{h_j}\right)\right] \\
&\quad = {}^{\circ}\left[\sum_{e':\{1,\ldots,q\}\to\{1,\ldots,\nu\}} \frac{1}{\nu^{q}}\cdot \right. \\
&\qquad\qquad \left. \sum_{f:\{1\}\to\{1,\ldots,\nu\}} \frac{1}{\nu} \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} w^{\Upsilon}_{\langle e'(1),\ldots,e'(q),f(1)\rangle}\left(\bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha^{\varepsilon_j}_{h_j}\right)\right],
\end{aligned}
$$

where

$$e(i) = \begin{cases} e'(i) & \text{if } i \in \{1, \ldots, q\}, \\ f(1) & \text{if } i = q+1. \end{cases}$$

It now remains to show that

$$\sum_{f:\{1\}\to\{1,\ldots,\nu\}} \frac{1}{\nu} \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} w^{\Upsilon}_{\langle e'(1),\ldots,e'(q),f(1)\rangle} \left(\bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha_{h_j}^{\varepsilon_j} \right) = w^{\Upsilon}_{\langle e'(1),\ldots,e'(q)\rangle}(\Theta) \tag{12}$$

for arbitrary $e' : \{1, \ldots, q\} \to \hat{\Upsilon}$. There are $\vec{c} \in {}^*\mathbb{D}_{2^q}$, $\vec{d} \in {}^*\mathbb{D}_{2^{q+1}}$ such that

$$w^{\Upsilon}_{\langle e'(1),\ldots,e'(q)\rangle} = w_{\vec{c}},$$
$$w^{\Upsilon}_{\langle e'(1),\ldots,e'(q),f(1)\rangle} = w_{\vec{d}}.$$

Given β_j an atom of L_{q+1}, there is a unique atom α_i of L_q and a unique $\varepsilon \in \{0, 1\}$ such that

$$\beta_j = \alpha_i^{\varepsilon}.$$

Thus, we can unambiguously write $d_j = c_i^{\varepsilon}$ for these i, ε. We then obtain

$$\begin{aligned} \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} w^{\Upsilon}_{\langle e'(1),\ldots,e'(q),f(1)\rangle} \left(\bigvee_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \bigwedge_{j=1}^{n} \alpha_{h_j}^{\varepsilon_j} \right) &= \sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \prod_{j=1}^{n} c_{h_j}^{\varepsilon_j} \\ &= \prod_{j=1}^{n} (c^0_{h_j} + c^1_{h_j}). \end{aligned} \tag{13}$$

Since by picking row $f(1)$ as the $q+1$'st row we partition the occurrences of the atom α_j of L_q obtained by picking rows $e'(1), \ldots, e'(q)$ into occurrences of the atoms α_j^1 and α_j^0 of L_{q+1}, and this is the only way in which we obtain these atoms, we must have $c_i^0 + c_i^1 = c_i$ for each $i \in \{1, \ldots, 2^q\}$. Thus (13) gives

$$\prod_{j=1}^{n} (c^0_{h_j} + c^1_{h_j}) = \prod_{j=1}^{n} c_{h_j} = w_{\langle e'(1),\ldots,e'(q)\rangle}(\Theta).$$

The equation (12) now follows.

It remains to show that Weak Irrelevance holds for ${}^\circ\nabla^L_\Upsilon$. Let $\vartheta(a_1, \ldots, a_m)$, $\varphi(a_{m+1}, \ldots, a_{m+n})$ be state descriptions of L having no constant or predicates in common. We can assume that $\vartheta \in QFSL^1$, $\varphi \in QFSL^2$, where $L^1 \cap L^2 = \emptyset$ and $L^1 \cup L^2 = L$.

Let α_i range over the atoms of L^1, β_j over the atoms of L^2. Then we obtain in L^1 and L^2, respectively,

$$\vartheta(a_1,\dots,a_m) = \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i),$$
$$\varphi(a_{m+1},\dots,a_{m+n}) = \bigwedge_{j=1}^{n} \beta_{g_j}(a_{m+j}).$$

Suppose that $L^1 = \{P_1,\dots,P_p\}$, $L^2 = \{P_{p+1},\dots,P_{p+r}\}$. Then we obtain in L

$$\vartheta(a_1,\dots,a_m) = \bigvee_{1\le s_1,\dots,s_m\le 2^r} \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \wedge \beta_{s_i}(a_i),$$
$$\varphi(a_{m+1},\dots,a_{m+n}) = \bigvee_{1\le t_1,\dots,t_n\le 2^p} \bigwedge_{j=1}^{n} \alpha_{t_j}(a_{m+j}) \wedge \beta_{g_j}(a_{m+j}),$$

and by ULi for ${}^\circ\nabla^L_\Upsilon$,

$$ {}^\circ\nabla^{L^1}_\Upsilon(\vartheta) = {}^\circ\nabla^L_\Upsilon\left(\bigvee_{1\le s_1,\dots,s_m\le 2^r} \bigwedge_{i=1}^{m} \alpha_{h_i} \wedge \beta_{s_i}\right), \tag{14}$$
$$ {}^\circ\nabla^{L^2}_\Upsilon(\varphi) = {}^\circ\nabla^L_\Upsilon\left(\bigvee_{1\le t_1,\dots,t_n\le 2^p} \bigwedge_{j=1}^{n} \alpha_{t_j} \wedge \beta_{g_j}\right). \tag{15}$$

Now for $\vartheta \wedge \varphi$, we obtain in L

$$\begin{aligned}
&{}^\circ\nabla^L_\Upsilon(\vartheta\wedge\varphi)\\
&\quad= {}^\circ\nabla^L_\Upsilon\left(\bigvee_{1\le s_1,\dots,s_m\le 2^r}\ \bigvee_{1\le t_1,\dots,t_n\le 2^p} \left(\bigwedge_{i=1}^{n} \alpha_{h_i}\wedge\beta_{s_i}\right) \wedge \left(\bigwedge_{j=1}^{n} \alpha_{t_j}\wedge\beta_{g_j}\right)\right)\\
&\quad= \sum_{1\le s_1,\dots,s_m\le 2^r}\ \sum_{1\le t_1,\dots,t_n\le 2^p} {}^\circ\left[\sum_{e:\{1,\dots,q\}\to\{1,\dots,\nu\}} \frac{1}{\nu^q}\cdot\right.\\
&\qquad\qquad \left. w^\Upsilon_{\langle e(1),\dots,e(q)\rangle}\left(\left(\bigwedge_{i=1}^{n} \alpha_{h_i}\wedge\beta_{s_i}\right) \wedge \left(\bigwedge_{j=1}^{n} \alpha_{t_j}\wedge\beta_{g_j}\right)\right)\right]\\
&\quad= \sum_{1\le s_1,\dots,s_m\le 2^r}\ \sum_{1\le t_1,\dots,t_n\le 2^p} {}^\circ\left[\sum_{e:\{1,\dots,q\}\to\{1,\dots,\nu\}} \frac{1}{\nu^q}\cdot w^\Upsilon_{\langle e(1),\dots,e(q)\rangle}\left(\bigwedge_{i=1}^{n} \alpha_{h_i}\wedge\beta_{s_i}\right)\right.\\
&\qquad\qquad \left.\cdot w^\Upsilon_{\langle e(1),\dots,e(q)\rangle}\left(\bigwedge_{j=1}^{n} \alpha_{t_j}\wedge\beta_{g_j}\right)\right],
\end{aligned}$$

by IP for $w^{\Upsilon}_{\langle e(1),\dots,e(q)\rangle}$,

$$= \left(\sum_{1\le s_1,\dots,s_m\le 2^r}^{\circ} \left[\sum_{e:\{1,\dots,q\}\to\{1,\dots,\nu\}} \frac{1}{\nu^q} \cdot w^{\Upsilon}_{\langle e(1),\dots,e(q)\rangle} \left(\bigwedge_{i=1}^{n} \alpha_{h_i} \wedge \beta_{s_i} \right) \right] \right)$$
$$\cdot \left(\sum_{1\le t_1,\dots,t_n\le 2^p}^{\circ} \left[\sum_{e:\{1,\dots,q\}\to\{1,\dots,\nu\}} \frac{1}{\nu^q} \cdot w^{\Upsilon}_{\langle e(1),\dots,e(q)\rangle} \left(\bigwedge_{j=1}^{n} \alpha_{t_j} \wedge \beta_{g_j} \right) \right] \right)$$
$$= \left(\sum_{1\le s_1,\dots,s_m\le 2^r} {}^{\circ}\nabla^L_{\Upsilon} \left(\bigwedge_{i=1}^{m} \alpha_{h_i} \wedge \beta_{s_i} \right) \right) \cdot \left(\sum_{1\le t_1,\dots,t_n\le 2^p} {}^{\circ}\nabla^L_{\Upsilon} \left(\bigwedge_{j=1}^{n} \alpha_{t_j} \wedge \beta_{h_j} \right) \right)$$
$$= {}^{\circ}\nabla^L_{\Upsilon}(\vartheta) \cdot {}^{\circ}\nabla^L_{\Upsilon}(\varphi),$$

by (14) and (15). ⊣

We are now set up to prove the main result of this section:

Theorem 9 (The Representation Theorem for ULi). *Let w be a probability function on $L = L_q$. Then w satisfies ULi if and only if there exists some normalized σ-additive measure ρ such that*

$$w = \int {}^{\circ}\nabla^L_{\Upsilon} \, d\rho(\Upsilon). \tag{16}$$

Proof: By Theorem 8, it is straightforward to see that any w in the form (16) satisfies ULi, as it is a convex combination of ULi functions.

For the other direction, suppose w satisfied ULi. Then there is an extension w^{L_ν} of w to L_ν and we obtain for $\Theta(a_1,\dots,a_n)$ a state description of L,

$$w(\Theta) = \sum_{\substack{\Phi(a_1,\dots,a_\nu) \\ \Phi\models\Theta}} w^{L_\nu}(\Phi), \tag{17}$$

where Φ ranges over the state descriptions of L_ν. For a state description $\Upsilon(P_1,\dots,P_\nu,a_1,\dots,a_\nu)$, let

$$\bar{\Upsilon} = \{\Upsilon(P_{\sigma(1)},\dots,P_{\sigma(\nu)},a_{\tau(1)},\dots,a_{\tau(\nu)} \,|\, \sigma,\tau \text{ are permutations of } \{1,\dots,\nu\}\}.$$

Note that the sets $\bar{\Upsilon}$ partition the set of state descriptions of L_ν. We can now write (17) as

$$w(\Theta) = \sum_{\bar{\Upsilon}} \sum_{\substack{\Phi\in\bar{\Upsilon} \\ \Phi\models\Theta}} w^{L_\nu}(\Phi)$$
$$= \sum_{\bar{\Upsilon}} \frac{|\{\Phi\in\bar{\Upsilon} \,|\, \Phi\models\Theta\}|}{|\bar{\Upsilon}|} w^{L_\nu}\left(\bigvee \bar{\Upsilon}\right),$$

as w^{L_ν} is clearly constant on $\bar{\Upsilon}$ since it satisfies Px (and Ex).

Now the ratio

$$\frac{|\{\Phi \in \bar{\Upsilon} \mid \Phi \models \Theta\}|}{|\bar{\Upsilon}|}$$

is equal to the probability that by randomly picking distinct predicates $P_{i_1}, \ldots, P_{i_q}$ and constants $a_{j_1}, \ldots, a_{j_n}$, we have that

$$\Upsilon \models \sigma\Theta(a_{j_1}, \ldots, a_{j_n}),$$

where σ is (an initial segment of) the permutation of predicates of L_ν with $\sigma(k) = i_k$ for $k \in \{1, \ldots, q\}$.

Note that with our definition of ∇^L_Υ, we allow the same row to be picked multiple times, so not all picks of rows represent a permutation of the predicates. Thus the difference between the probabilities given by ∇^L_Υ and the above ratio is the difference between picking rows of Υ with and without replacement. However, since the probability of picking the same row twice is infinitesimal, it will disappear when taking standard parts.

Thus we obtain

$$^\circ\left(\frac{|\{\Phi \in \bar{\Upsilon} \mid \Phi \models \Theta\}|}{|\bar{\Upsilon}|}\right) = {}^\circ\nabla^L_\Upsilon(\Theta).$$

Now taking μ to be the measure on the $\bar{\Upsilon}$ given by w^{L_ν}, we obtain

$$\sum_{\bar{\Upsilon}} \frac{|\{\Phi \in \bar{\Upsilon} \mid \Phi \models \Theta\}|}{|\bar{\Upsilon}|} w^{L_\nu}\left(\bigvee \bar{\Upsilon}\right) = \int \frac{|\{\Phi \in \bar{\Upsilon} \mid \Phi \models \Theta\}|}{|\bar{\Upsilon}|}\, d\mu(\bar{\Upsilon}).$$

Taking standard parts, we obtain

$$\begin{aligned} {}^\circ\!\int \frac{|\{\Phi \in \bar{\Upsilon} \mid \Phi \models \Theta\}|}{|\bar{\Upsilon}|}\, d\mu(\bar{\Upsilon}) &= \int {}^\circ\left(\frac{|\{\Phi \in \bar{\Upsilon} \mid \Phi \models \Theta\}|}{|\bar{\Upsilon}|}\right) d\rho(\bar{\Upsilon}) \\ &= \int {}^\circ\nabla^L_\Upsilon\, d\rho(\bar{\Upsilon}), \end{aligned}$$

where ρ is the Loeb measure given by the nonstandard measure μ. $\dashv$

Since ${}^\circ\nabla^L_\Upsilon$ satisfies WIP we obtain the following theorem.

Theorem 10. *The ${}^\circ\nabla^L_\Upsilon$ are the only functions satisfying ULi with WIP.*

Proof: We follow essentially the proof for the analogous theorem for Atom Exchangeability.[10]

Let w be a probability function satisfying ULi with WIP. Let $\vartheta \in QFSL$. Extend w to w' on some language L' large enough so that we can permute the predicates and constants in ϑ to obtain ϑ' with no predicates nor constants in common with ϑ. We can achieve this by picking w' on L' in the same ULi family as w, giving $w' \restriction SL = w$ and guaranteeing WIP for w'. By Px for w' we then have $w'(\vartheta) = w'(\vartheta')$. Now we clearly obtain

$$\begin{aligned}
0 &= 2(w'(\vartheta \wedge \vartheta') - w'(\vartheta) \cdot w'(\vartheta')) \\
&= \int {}^\circ\nabla_\Psi^{L'}(\vartheta \wedge \vartheta')\, d\mu(\Psi) - 2\int {}^\circ\nabla_\Psi^{L'}(\vartheta)\, d\mu(\Psi) \cdot \int {}^\circ\nabla_\Phi^{L'}(\vartheta')\, d\mu(\Phi) \\
&\qquad + \int {}^\circ\nabla_\Phi^{L'}(\vartheta \wedge \vartheta')\, d\mu(\Phi) \\
&= \int\int \left({}^\circ\nabla_\Psi^{L'}(\vartheta)^2 - 2{}^\circ\nabla_\Psi^{L'}(\vartheta) \cdot {}^\circ\nabla_\Phi^{L'}(\vartheta) + {}^\circ\nabla_\Phi^{L'}(\vartheta)^2\right) d\mu(\Psi)\, d\mu(\Phi) \\
&= \int\int \left({}^\circ\nabla_\Psi^{L'}(\vartheta) - {}^\circ\nabla_\Phi^{L'}(\vartheta)\right)^2 d\mu(\Psi)\, d\mu(\Phi),
\end{aligned}$$

using the Representation Theorem. Certainly, since the function under the integral is non-negative, there must be a measure 1 set such that ${}^\circ\nabla_\Psi^{L'}$ is constant on this set for each $\vartheta \in QFSL$, giving $w' = {}^\circ\nabla_\Psi^{L'}$ for any Ψ in this set. Since $w' \restriction SL = w$, i.e. $w = {}^\circ\nabla_\Psi^{L'} \restriction SL$, marginalizing w' to L yields $w = {}^\circ\nabla_\Psi^{L}$, as required. $\dashv$

With Theorem 9, we have shown that the building blocks for probability functions satisfying Unary Language Invariance all satisfy Weak Irrelevance, and that in fact these are the only ones that satisfy this principle. This is analogous to the situation with Atom Exchangeability and its generalization to Polyadic Pure Inductive Logic, Spectrum Exchangeability, see [12, Chapter 32].

5 A General Representation Theorem

In the case of Atom Exchangeability, Ax, we have a theorem stating that each w satisfying Ax can be represented as a difference of scaled ULi functions with Ax (see e.g. [12, chapter 33]). In this section, we will prove an analogous result for Px. For the remainder of this section we assume that $L = L_q$ for some $q \in \mathbb{N}$.

[10] The principle of Atom Exchangeability is a strengthening of Px stating that a probability function w should be invariant under permutations of the *atoms* of the language L (see e.g. [12, Chapter 14]). For the purposes of this paper however it is not necessary to know anything more about Atom Exchangeability.

Definition 11:
Let $\vec{c} \in \mathbb{D}_{2^q}$. Let Σ be the set of all permutations of atoms of L that are induced by Px. Define the probability function $y_{\vec{c}}$ on $QFSL$ by

$$y_{\vec{c}}(\Theta(a_1, \dots, a_n)) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} w_{\sigma\vec{c}}(\Theta(a_1, \dots, a_n))$$

for state descriptions $\Theta(a_1, \dots, a_n)$ of L.

Note that by definition, $y_{\vec{c}}$ satisfies Px. By a straightforward argument we obtain the following variation on de Finetti's Theorem:

Theorem 12. *Let w be a probability function on SL satisfying Px. Then there exists a normalized, σ-additive measure μ on the Borel sets of $\mathbb{D}_{2^q}$ such that*

$$w\left(\bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) = \int_{\mathbb{D}_{2^q}} y_{\vec{c}}\left(\bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) d\mu(\vec{c}). \tag{18}$$

Conversely, given such a measure μ, the function w defined by (18) *satisfies Px.*

The key to obtaining the desired General Representation Theorem will therefore involve finding a uniform representation of the building blocks $y_{\vec{c}}$ in terms of a difference of ULi functions. The ${}^\circ\nabla^L_\Upsilon$ functions used for this proof will have a specific characterization that deserves a slightly different notation. Since at this point, we will be working in the usual standard universe again, we will drop the standard part symbol $^\circ$ from the notation and assume that all ∇^L_Υ from now on are given in their standard form.

Recalling the definition of ∇^L_Υ note that for fixed $e : \{1, \dots, q\} \to \{1, \dots, \nu\}$, the function $w^\Upsilon_{\langle e(1),\dots,e(q)\rangle}$ is given by the $q \times \nu$ - matrix with the i'th row identical to the $e(i)$'th row of Υ. Also, since with $w^\Upsilon_{\langle e(1),\dots,e(q)\rangle}$ we also have all the $w^\Upsilon_{\langle \sigma(e(1)),\dots,\sigma(e(q))\rangle}$ for σ ranging over the permutations of the predicates of L occurring in ∇^L_Υ, we see that this function is a convex combination of functions of the form $y_{\vec{c}}$.

We can now arrange ∇^L_Υ to contain a copy of $y_{\vec{c}}$ for a given $\vec{c} \in \mathbb{D}_{2^q}$ as follows: Let Φ be the state description represented by the matrix

$$\begin{pmatrix} | & & | & | & & | & & | & & | \\ \alpha_1 & \cdots & \alpha_1 & \alpha_2 & \cdots & \alpha_2 & \cdots & \alpha_{2^q} & \cdots & \alpha_{2^q} \\ | & & | & | & & | & & | & & | \end{pmatrix},$$

where α_i occurs $[c_i \cdot \nu]$ times. Now let $\mathfrak{p}_1, \dots, \mathfrak{p}_q \geq 0$ be such that $\sum_{i=1}^q \mathfrak{p}_i = 1$ and let Υ be the $\nu \times \nu$ - matrix containing $[\mathfrak{p}_i \cdot \nu]$ copies of the i'th row of Φ, for each i, and fill the remaining rows with arbitrary copies of rows from Φ. Then ∇^L_Υ certainly contains a copy of $y_{\vec{c}}$.

With this in mind, we can modify the notation of ∇^L_Υ to

$$\vec{\mathfrak{p}}\nabla^L_\Upsilon$$

for $\vec{\mathfrak{p}} = \langle \mathfrak{p}_1, \ldots, \mathfrak{p}_q \rangle$ to indicate that Υ contains only q distinct rows, occurring with the frequency given by $\vec{\mathfrak{p}}$. We will write $\vec{\mathfrak{p}}\nabla^L_{\Upsilon(\vec{c})}$ to indicate that Υ arises from $\vec{c} \in \mathbb{D}_{2^q}$ in this manner.

We can represent $\vec{\mathfrak{p}}\nabla^L_{\Upsilon(\vec{c})}$ in terms of $y_{\vec{c}}$ as follows. Let $K = \{\vec{n} \in \mathbb{N}^q \mid \sum_{i=1}^q n_i = q\}$, so $\vec{n} \in K$ represents the choices of picking rows from Υ. Then we obtain the representation

$$\vec{\mathfrak{p}}\nabla^L_{\Upsilon(\vec{c})} = \sum_{\vec{n} \in K} \prod_{i=1}^{q} \mathfrak{p}_i^{n_i} (n_1, \ldots, n_q)!\, y_{\vec{c}_{\vec{n}}}, \tag{19}$$

where $\vec{c}_{\vec{n}}$ results from picking rows according to $\vec{n}$ and (as standard)

$$(n_1, \ldots, n_q)! = \frac{(n_1 + n_2 + \ldots + n_q)!}{n_1!\, n_2! \ldots n_q!} = \binom{q}{n_1, \ldots, n_q}.$$

Note that we need this multinomial coefficient here since $\vec{\mathfrak{p}}\nabla^L_{\Upsilon(\vec{c})}$ is in fact a sum of $w_{\vec{e}}$, and although each of the $w_{\vec{e}}$ occurring in $y_{\vec{c}}$ occurs, the normalizing constant exists only implicitly in $\vec{\mathfrak{p}}\nabla^L_{\Upsilon(\vec{c})}$. With this notation in mind, we can prove the first step needed to show the desired theorem.

Lemma 13. *Let $\vec{c} \in \mathbb{D}_{2^q}$. Then there exist $\lambda \geq 0$ and probability functions w_1, w_2 satisfying ULi such that*

$$y_{\vec{c}} = (1 + \lambda) w_1 - \lambda w_2.$$

Proof: Fix $\vec{c} \in \mathbb{D}_{2^q}$. As demonstrated in the discussion above, we can easily find ∇^L_Υ with $y_{\vec{c}}$ occurring in it, amongst other instances of $y_{\vec{e}}$. Thus, the problem reduces to finding a way to remove all of these other instances via ULi functions.

To this end, suppose that for each $\vec{m} \in K$ we have $\vec{\mathfrak{p}}_{\vec{m}}\nabla^L_{\Upsilon(\vec{c})}$ such that Υ is the state description obtained from $w_{\vec{c}}$ by the method discussed above. Then, since the representations of the form (19) of these functions only differ in the coefficients of the $y_{\vec{e}}$ occurring we obtain the equation

$$\begin{pmatrix} \vdots \\ \vec{\mathfrak{p}}_{\vec{m}}\nabla^L_{\Upsilon(\vec{c})} \\ \vdots \end{pmatrix} = A \cdot \begin{pmatrix} \vdots \\ (m_1, \ldots, m_q)!\, y_{\vec{c}_{\vec{m}}} \\ \vdots \end{pmatrix}, \tag{20}$$

where A is the $K \times K$-matrix with entry $\langle \vec{m}, \vec{n} \rangle$ being $\prod_{k=1}^q \mathfrak{p}_{\vec{m},k}^{n_k}$. It suffices now to show that we can pick the $\vec{\mathfrak{p}}_{\vec{m}}$ such that A is regular. For suppose this is the case. Then we obtain from (20) the equation

$$A^{-1} \begin{pmatrix} \vdots \\ \vec{\mathfrak{p}}_{\vec{m}}\nabla^L_{\Upsilon(\vec{c})} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ (m_1, \ldots, m_q)!\, y_{\vec{c}_{\vec{m}}} \\ \vdots \end{pmatrix}.$$

Suppose $A^{-1} = (b_{\vec{n},\vec{m}})_{\vec{n},\vec{m}\in K}$. Then for $\vec{n} = \langle 1, 1, \dots, 1\rangle$ we obtain

$$y_{\vec{c}} = \frac{1}{(n_1, \dots, n_q)!} \sum_{\vec{m}\in K} b_{\vec{n},\vec{m}} \vec{\mathfrak{p}}_{\vec{m}} \nabla^L_{\Upsilon(\vec{c})} = \frac{1}{q!} \sum_{\vec{m}\in K} b_{\vec{n},\vec{m}} \vec{\mathfrak{p}}_{\vec{m}} \nabla^L_{\Upsilon(\vec{c})},$$

and by collecting the functions with positive coefficients in the linear combination on the right-hand side, we obtain constants $\gamma, \lambda \geq 0$, *independent of* $\vec{c}$, such that[11]

$$\frac{1}{q!} \sum_{\vec{m}\in K} b_{k,\vec{m}} \vec{\mathfrak{p}}_{\vec{m}} \nabla^L_{\Upsilon(\vec{c})} = \gamma w_1 - \lambda w_2,$$

with w_1, w_2 convex combinations of ULi functions. Since this gives the probability function $y_{\vec{c}}$, we must have

$$1 = y_{\vec{c}}(\top) = \gamma w_1(\top) - \lambda w_2(\top) = \gamma - \lambda,$$

and thus $\gamma = 1 + \lambda$.

It remains to show that the $\vec{\mathfrak{p}}_{\vec{m}}$ can be chosen such that A is regular. For this, we will show the following by induction on j:
Let $1 \leq i_1 < i_2 < \cdots < i_j \leq r$ and let $A_{\langle i_1,\dots,i_j\rangle}$ be the $j \times j$ sub-matrix of A obtained by taking the $i_1, \dots, i_j$'th rows and columns of A. Then there is a choice of the $\vec{\mathfrak{p}}_{\vec{m}_k}$, $k = i_1, \dots, i_j$ such that $A_{\langle i_1,\dots,i_j\rangle}$ is regular.

For $j = 1$, this is trivial. Suppose $j = n+1$ for some $n \geq 1$ and consider $A_{\langle i_1,\dots,i_j\rangle}$. For a given $\vec{m} \in K$, the polynomial $\prod_{j=1}^q x_j^{m_j}$ takes its maximum value on $\mathbb{D}_{2^q}$ at $x_j = m_j/q$. Fix an enumeration of K. There exists $\vec{m}_{i_k} = \langle m_{i_k,1}, \dots, m_{i_k,q}\rangle$ such that

$$\prod_{s=1}^q \left(\frac{m_{i_k,s}}{q}\right)^{m_{i_k,s}} > \prod_{s=1}^q \left(\frac{m_{i_k,s}}{q}\right)^{m_{i_j,s}}$$

for all $j \neq k$. For if not, then

$$\prod_{s=1}^q \left(\frac{m_{i_k,s}}{q}\right)^{m_{i_k,s}} \leq \prod_{s=1}^q \left(\frac{m_{i_k,s}}{q}\right)^{m_{i_j,s}} < \prod_{s=1}^q \left(\frac{m_{i_j,s}}{q}\right)^{m_{i_j,s}}$$

for some $j \neq k$, and continuing in this way we arrive at a contradiction.

By the inductive hypothesis, there exists a choice of the $\vec{\mathfrak{p}}_{\vec{m}_s}$, $s \in \{i_1, \dots, i_j\} \setminus \{i_k\}$ such that the sub-matrix $A_{\langle i_1,\dots,i_{k-1},i_{k+1},\dots,i_j\rangle}$ is regular. Thinking of the $\mathfrak{p}_{\vec{m}_{i_k},s}$ for the moment as unknowns we obtain for the determinant of $A_{\langle i_1,\dots,i_j\rangle}$ an expression of the form

$$\det(A_{\langle i_1,\dots,i_j\rangle}) = \\ \pm \prod_{s=1}^q \mathfrak{p}_{\vec{m}_{i_k},s}^{m_{i_k,s}} \cdot \det(A_{\langle i_1,\dots,i_{k-1},i_{k+1},\dots,i_j\rangle}) + \sum_{t\in\{i_1,\dots,i_j\}\setminus\{i_k\}} \prod_{s=1}^q \mathfrak{p}_{\vec{m}_{i_k},s}^{m_{t,s}} \cdot (\pm \det(A_t)), \quad (21)$$

[11] Note that we can safely assume $\lambda \neq 0$, since if $\lambda = 0$, then the $y_{\vec{c}}$ in question would already satisfy ULi, and therefore already has the desired representation by the Representation Theorem for ULi. We also trivially have $\gamma \neq 0$, since $y_{\vec{c}}$ is a probability function for any $\vec{c} \in \mathbb{D}_{2^q}$.

(for some choices of $\pm$) where the A_t are the corresponding sub-matrices of $A_{\langle i_1,\ldots,i_j\rangle}$. Now picking $\mathfrak{p}_{\vec{m}_{i_k},s} = (m_{i_k,s}/q)^g$ for large enough $g > 0$, the term

$$\prod_{s=1}^{q} \mathfrak{p}_{\vec{m}_{i_k},s}^{m_{i_k,s}} \cdot \det(A_{\langle i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_j\rangle})$$

becomes the dominant term of (21), giving that $\det(A_{\langle i_1,\ldots,i_j\rangle}) \neq 0$, as certainly $\prod_{s=1}^{q} \mathfrak{p}_{\vec{m}_{i_k},s}^{n_{i_k,s}} > 0$ and $\det(A_{\langle i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_j\rangle}) \neq 0$ by the inductive hypothesis.

Note that using this procedure we in general obtain $\vec{\mathfrak{p}}_{\vec{m}}$ with entries $\mathfrak{p}_{\vec{m}_i,j}$ not summing to 1. In that case, we can pick $\vec{\mathfrak{p}'}_{\vec{m}}$ such that

$$\mathfrak{p}'_{\vec{m}_i,j} = \frac{\mathfrak{p}_{\vec{m}_i,j}}{\sum_{s=1}^{q} \mathfrak{p}_{\vec{m}_i,s}}$$

for each $\vec{m} \in K$. Then the matrix A' with entries $\prod_{s=1}^{q} \mathfrak{p}'^{n_{j,s}}_{\vec{m}_i,s}$ is regular just if A is, and the $\vec{\mathfrak{p}'}_{\vec{m}}$ have the desired properties. $\dashv$

Using this lemma, we can now prove the desired theorem.

Theorem 14 (The General Representation Theorem for Px). *Let w be a probability function on SL satisfying Px. Then there exist $\lambda \geq 0$ and probability functions w_1, w_2 satisfying ULi such that*

$$w = (1+\lambda)w_1 - \lambda w_2.$$

Proof: Let w be a probability function on SL satisfying Px. By the Representation Theorem for Px, we have that w has a representation

$$w = \int_{\mathbb{D}_{2^q}} y_{\vec{c}} \, d\mu(\vec{c}) \tag{22}$$

for some measure μ, and by Lemma 13, we have, for a fixed $\lambda \geq 0$, a representation

$$y_{\vec{c}} = (1+\lambda)w_{1_{\vec{c}}} - \lambda w_{2_{\vec{c}}}$$

for each $\vec{c} \in \mathbb{D}_{2^q}$. Now applying this to the representation (22), we obtain

$$\begin{aligned} w &= \int_{\mathbb{D}_{2^q}} (1+\lambda)w_{1_{\vec{c}}} - \lambda w_{2_{\vec{c}}} \, d\mu(\vec{c}) \\ &= \int_{\mathbb{D}_{2^q}} (1+\lambda)w_{1_{\vec{c}}} \, d\mu(\vec{c}) - \int_{\mathbb{D}_{2^q}} \lambda w_{2_{\vec{c}}} \, d\mu(\vec{c}) \\ &= (1+\lambda)w_1 - \lambda w_2, \end{aligned}$$

for

$$w_1 = \int_{\mathbb{D}_{2^q}} w_{1_{\vec{c}}} \, d\mu(\vec{c}), \qquad w_2 = \int_{\mathbb{D}_{2^q}} w_{2_{\vec{c}}} \, d\mu(\vec{c}),$$

as required. $\dashv$

Theorem 14 might initially suggest that if our agent chooses a probability function w on a language $\mathcal{L}$ then she can express w in the form $(1+\lambda)w_1 + \lambda w_2$, where w_1, w_2 satisfy ULi, and hence extend w to a larger language by using this form with w_1, w_2 extended to this larger language. The flaw in this argument is that for ϑ from a larger language there is no longer any guarantee that

$$(1+\lambda)w_1(\vartheta) \geq \lambda w_2(\vartheta).$$

In other words such an attempt to extend w can (in fact has to) lead to 'negative probabilities'.

Again, as with Theorem 9, an analogous General Representation Theorem to Theorem 14 has been proved for Atom Exchangeability and its generalization to Polyadic Pure Inductive Logic, Spectrum Exchangeability, see [12, Chapter 34].

6 Conclusion

The three main results in this paper are the Representation Theorems 5 and 9 for ULi + IP and ULi respectively and the General Representation Theorem 14. In the process we also obtained a complete characterization of the probability functions satisfying ULi with Weak Irrelevance, Theorem 10. Mathematically such representation theorems are valuable because they tell us how probability functions satisfying, for example ULi, are made up of simple building block functions satisfying ULi and as a result it is often the case that to prove some property holds of all probability functions satisfying ULi it is enough to show it for these simple building blocks, for which we usually have a much clearer grasp. Whilst the mathematics may be somewhat technical at times there are numerous examples where this has led to results which are philosophically interesting, particular examples of this being Humburg's use of de Finetti's Representation Theorem to prove that Ex implies the Principle of Instantial Relevance, see [9], and the recent use of the polyadic version of de Finetti's Theorem to refine the Counterpart Principle of Analogy, see [12, Chapter 22].

In particular in this paper we have obtained results characterizing the probability functions satisfying Px and relating this principle to ULi, IP and WIP, all four principles which are directly accessible to philosophical consideration. Indeed it seems to us hard not to grant Px the same degree of acceptance as Ex commonly now enjoys within the context of Pure Inductive Logic, and in which case investigating its properties and relationships with other purportedly rational principles is central.

Throughout this paper we have worked in the conventional *Unary* Pure Inductive Logic. Over the past decade however there has been a rapid development of *Polyadic* Pure Inductive Logic (again see [12]) and we anticipate that the Representation Theorem for ULi functions can be extended to the polyadic case, using the same methods as demonstrated above. A classification for probability functions on polyadic languages satisfying Language Invariance would give rise to the question whether we can find a corresponding General Representation Theorem for the polyadic case as well.

7 Acknowledgement

We would like to thank the referees for their suggestions for clarifying the original version of this paper.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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, Harlow, 1995, pp151-177.

[5] de Finetti, B., *Theory of Probability, Volume 1*, Wiley, New York, 1974.

[6] Gaifman, H., Applications of de Finetti's Theorem to Inductive Logic, in *Studies in Inductive Logic and Probability*, Volume I, Eds. R.Carnap & R.C.Jeffrey, University of California Press, 1971, pp235-251.

[7] Gaifman, H., Concerning measures on first order calculi, *Israel Journal of Mathematics*, 1964, **2**:1-18.

[8] Hill, M.J., Paris, J.B. & Wilmers, G.M., Some observations on induction in predicate probabilistic reasoning, *Journal of Philosophical Logic*, 2002, **31**:43-75.

[9] Humburg, J., The Principle of Instantial Relevance, *Studies in Inductive Logic and Probability, Volume I*, Eds. R.Carnap & R.C.Jeffrey, University of California Press, Berkeley and Los Angeles, 1971, pp225-233.

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

[11] Paris, J.B., Pure Inductive Logic, in *The Continuum Companion to Philosophical Logic*, Eds. L.Horsten, R.Pettigrew, Continuum International Publishing Group, London, 2011, pp428-449.

[12] Paris, J.B. and Vencovská, A., *Pure Inductive Logic*, to appear in the Association of Symbolic Logic series Perspectives in Mathematical Logic, Cambridge University Press, 2013.

[13] Paris, J.B. and Vencovská, A., A Note on Irrelevance in Inductive Logic, *Journal of Philosophical Logic*, 2011, **40**:357-370.

[14] Paris, J.B. & Vencovská, A., A General Representation Theorem for Probability Functions Satisfying Spectrum Exchangeability, in *CiE 2009*, Eds. K.Ambros-Spies, B.Löwe & W.Merkle, Springer LNCS 5635, pp379-388.