The Theory of Spectrum Exchangeability

Howarth, E. and Paris, J.B.

2015

MIMS EPrint: **2015.71**

ISSN 1749-9097

The Theory of Spectrum Exchangeability*

E.Howarth† and J.B.Paris
School of Mathematics
The University of Manchester
Manchester M13 9PL

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

August 3, 2015

Abstract

Spectrum Exchangeability, Sx, is an irrelevance principle of Pure Inductive Logic, and arguably the most natural (but not the only) extension of Atom Exchangeability to polyadic languages. It has been shown[1] that all probability functions which satisfy Sx are comprised of a mixture of two essential types of probability functions; heterogeneous and homogeneous functions. We determine the theory of Spectrum Exchangeability, which for a fixed language L is the set of sentences of L which must be assigned probability 1 by every probability function satisfying Sx, by examining separately the theories of heterogeneity and homogeneity. We find that the theory of Sx is equal to the theory of finite structures, i.e. those sentences true in all finite structures for L, and it emerges that Sx is inconsistent with the principle of Super-Regularity (Universal Certainty). As a further consequence we are able to characterize those probability functions which satisfy Sx and the Finite Values Property.

*This paper is published as open access in the *Review of Symbolic Logic,* **8**(01):108-130, March 2015.

†Supported by a UK Engineering and Physical Sciences Research Council (EPSRC) Studentship.

[1]See [6], [8], [9], [10], [15]

Key words: Spectrum Exchangeability, Finite Values Property, Inductive Logic, Logical Probability, Rationality, Uncertain Reasoning.

Introduction

The framework of this paper is that of Pure Inductive Logic, PIL, as presented in [12], [15], according to which PIL is the mathematical formalization of

> "*assigning logical, as opposed to statistical, probabilities by attempting to formulate the underlying notions, such as symmetry, irrelevance, relevance on which they appear to depend*".

Within PIL, reasoning is modelled by the choice of a single probability function, from all possible probability functions defined on the sentences of a first-order predicate language L, so that an agent's 'belief' that a given sentence is true is represented by the value in the interval $[0, 1]$ assigned to that sentence by the agent's chosen function. The agent is assumed to have no interpretation of the language; so that its choice of probability function must be based on logical considerations alone.

We consider how a supposedly rational agent inhabiting a structure for L, but having no prior knowledge concerning which such structure, should assign probabilities $w(\theta)$ to the sentences θ of L. Or putting it another way, how does the requirement of rationality restrict the agent's choice of probability function?

A standard procedure for investigating this question is to propose purportedly rational principles which one may feel the agent should observe, and then investigate their consequences. One approach is to examine the resulting *theory*; the set of sentences of L which must be assigned probability 1 by any probability function which satisfies the chosen principle(s). If this set $Th(\mathcal{P})$ can be identified for a particular set of principles $\mathcal{P}$, this gives a kind of 'creed' according to $\mathcal{P}$, a set of sentences which must be accepted with certainty by any agent who adopts $\mathcal{P}$. By the definition of a probability function, $Th(\mathcal{P})$ must contain all tautologies, but where it additionally contains non-tautologous sentences, this surely says something interesting about $\mathcal{P}$, which may give a new perspective on the extent to which $\mathcal{P}$ is a 'good' choice of rational principles.

In this paper we consider separately the theories of the two essential types of probability functions satisfying the rational principle of Spectrum Exchangeability, Sx, namely heterogeneous and homogeneous functions. We show that the theory of heterogeneity is equal to the theory of finite structures for L, i.e. those sentences true in all finite structures for L, which in turn is equal to the theory of Sx. It follows as a corollary that the principle of Sx is incompatible with that of Super-Regularity (that all non-contradictory sentences should be assigned non-zero probability).

Several other principles are known to have non-trivial theories, in particular the Invariance Principle INV (see [15]) and, for purely unary languages Johnson's Sufficient Postulate JSP (see [5] for an explicit description of $Th(JSP)$). A feature of such principles seems to be that they are rather powerful, suggesting, rather reasonably, that this power comes at the price of the agent accepting a non-tautological 'creed'.

A further consequence of our investigations into the theory of Sx is that we are able to characterize those probability functions satisfying Sx which also satisfy the Finite Values Property, meaning that for a fixed finite number of constant symbols, the probability function only takes finitely many different values on sentences involving just those constant symbols.

Context and Notation

The context described in the introduction, and based on that considered in [15], consists of a first order language L containing finitely many (possibly polyadic) relation symbols $\{R_1, R_2, \ldots, R_q\}$, with variables x_i and constant symbols a_i for $i \in \{1, 2, 3, \ldots\} = \mathbb{N}^+$. Equality and function symbols are assumed to be absent. For convenience of notation, alternative symbols may be used for constants and variables, such as $b_1, \ldots, b_m$ to represent a sequence of m unspecified constant symbols $a_{i_1}, \ldots, a_{i_m}$, assumed to be distinct unless stated otherwise. Similarly y_j, z_k etc. may be used for variables.

Let SL denote the set of first order sentences of L and $QFSL$ denote those sentences of SL which are quantifier free. Similarly, let $(QF)FL$ denote the (quantifier free) formulae of L. Where a sentence is denoted $\theta(b_1, \ldots, b_m)$, this expresses that θ mentions only constants from among $b_1, \ldots, b_m$, but not necessarily all (or any) of these.

Let $\mathcal{T}L$ denote the set of structures for L with universe $\{a_1, a_2, a_3, \ldots\}$, where the symbol a_i is interpreted as the individual a_i (and no distinction is made between the two). Notice that by the Löwenheim-Skolem Theorems if $\Gamma \subseteq SL$ and infinitely many constants are not mentioned by the sentences of Γ then Γ has a model just if it has a model in $\mathcal{T}L$. With this limitation on Γ then $\mathcal{T}L$ is *complete*, in the sense that if a sentence ψ is true in all models of Γ in $\mathcal{T}L$ then it is true in *all* models of Γ, and hence formally provable from Γ (indeed from some finite subset of Γ). This is a property of $\mathcal{T}L$ that we will need later.

The central question of PIL is how a supposedly rational agent inhabiting a structure M in $\mathcal{T}L$, but having no prior knowledge concerning which such structure, should assign probabilities $w(\theta)$ to the sentences $\theta \in SL$. Probability functions defined on SL are used to model the agent's belief, see for example [3], so in these terms the essential question is to what extent the requirement of rationality limits the agent's choice of probability function.

A function $w : SL \to [0,1]$ is a *probability function* on SL just if it satisfies that for all $\theta, \phi, \exists x\, \psi(x) \in SL$:

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

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

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

All the standard properties of probability functions readily follow from (P1–3), see for example [11, Proposition 2.1], [15].

Conditioning is used to model the process of the agent's learning, or imagining that it has learnt, that some sentence is true in M. For a probability function w on SL and a fixed $\phi \in SL$ the *conditional probability* function of w given ϕ is defined to be probability function $w(\cdot \mid \phi)$ such that for $\theta \in SL$

$$w(\theta \mid \phi) \cdot w(\phi) = w(\theta \wedge \phi).$$

In particular then, when $w(\phi) > 0$,

$$w(\theta \mid \phi) = \frac{w(\theta \wedge \phi)}{w(\phi)}.$$

The following principle of rationality is based on the symmetry between the constant symbols of L.

Constant Exchangeability, Ex.

A probability function w on SL satisfies Constant Exchangeability if, for any permutation σ of $1, 2, \ldots$ and $\theta(a_1, \ldots, a_m) \in SL$,

$$w(\theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)})) = w(\theta(a_1, \ldots, a_m)). \tag{1}$$

The argument for this principle is that there is complete symmetry between the constants, and hence between $\theta(a_1, \ldots, a_m)$ and $\theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)})$, so it would be irrational to assign these two sentences different probabilities. All probability functions considered in this paper are assumed to satisfy Ex.

A *state description* for $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ is a quantifier free sentence $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_m})$ of the form

$$\bigwedge_{k=1}^{q} \bigwedge_{\vec{b}} \pm R_k(b_1, b_2, \ldots, b_{r_k})$$

where r_k is the arity of the relation symbol R_k and the $\vec{b} = \langle b_1, b_2, \ldots, b_{r_k} \rangle$ range over all possible tuples from $\{a_{i_1}, \ldots, a_{i_m}\}^{r_k}$. Here, $+R_k(a_i)$ stands for $R_k(a_i)$, while $-R_k(a_i)$ stands for $\neg R_k(a_i)$. A formula of this form for distinct variables $x_{i_1}, \ldots, x_{i_m}$ is known as a *state formula* for $x_{i_1}, \ldots, x_{i_m}$. By convention, state descriptions for zero constants and state formulae for zero variables are taken to be equivalent to some fixed tautology, denoted $\top$, mentioning no constants. Upper case Greek letters will always be used to denote state descriptions or state formulae.

By the Disjunctive Normal Form Theorem, every $\phi(a_{i_1}, a_{i_2}, \ldots, a_{i_m}) \in QFSL$ is logically equivalent to a disjunction of (necessarily pairwise disjoint) state descriptions, from which it follows that the probability of ϕ is the sum of the probabilities of these state descriptions. Furthermore, by Gaifman's Theorem [4], a probability function is completely determined on the whole of SL, not just on $QFSL$, by its values on state descriptions.

A restriction of a state description $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_m})$ to a subset $\{a_{i_{s_1}}, \ldots, a_{i_{s_u}}\}$ of $\{a_{i_1}, \ldots, a_{i_m}\}$, i.e. the conjunction of those conjuncts in Θ which refer only to constants from among $\{a_{i_{s_1}}, \ldots, a_{i_{s_u}}\}$, will be denoted $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_m})[a_{i_{s_1}}, \ldots, a_{i_{s_u}}]$, or just $\Theta[a_{i_{s_1}}, \ldots, a_{i_{s_u}}]$ if the tuple $\langle a_{i_1}, a_{i_2}, \ldots, a_{i_m} \rangle$ is clear from the context. For example if L has just a single binary relation symbol R and $\Theta(a_1, a_2, a_3)$ is the conjunction of

$$\begin{array}{ccc} R(a_1, a_1) & R(a_1, a_2) & R(a_1, a_3) \\ \neg R(a_2, a_1) & R(a_2, a_2) & \neg R(a_2, a_3) \\ \neg R(a_3, a_1) & \neg R(a_3, a_2) & R(a_3, a_3) \end{array}$$

then $\Theta(a_1,a_2,a_3)[a_1,a_3]$ is the conjunction of

$$\begin{array}{cc} R(a_1,a_1) & R(a_1,a_3) \\ \neg R(a_3,a_1) & R(a_3,a_3). \end{array}$$

A state description $\Psi(a_{i_1},\dots,a_{i_m},a_{i_{m+1}})$ *extends* $\Theta(a_{i_1},\dots,a_{i_m})$, equivalently $\Psi \models \Theta$, if the restriction of Ψ to $a_{i_1},\dots,a_{i_m}$ is logically equivalent to Θ: $\Psi[a_{i_1},\dots,a_{i_m}] \equiv \Theta(a_{i_1},\dots,a_{i_m})$. The same notation is used for restrictions and extensions of state formulae.

For a fixed language L, any $n \geq m \geq 0$ and any state description $\Theta(a_{i_1},\dots,a_{i_m})$ of L, the number of state descriptions $\Phi(a_{i_1},\dots,a_{i_n})$ of L extending Θ depends only on m and n (and L, which we leave implicit). This number will be denoted $SD(m,n)$, while the total number of state descriptions in L for n constants is denoted $SD(n)\ (= SD(0,n))$.

Suppose that a state description $\Theta(b_1,b_2,\dots,b_m)$ is such that for some b_i,b_j

$$\Theta \models R(b_{k_1},\dots,b_{k_u},b_i,b_{k_{u+2}},\dots,b_{k_r}) \leftrightarrow R(b_{k_1},\dots,b_{k_u},b_j,b_{k_{u+2}},\dots,b_{k_r})$$

for any r-ary relation symbol from L and any (not necessarily distinct) $b_{k_1},\dots,b_{k_u}$, $b_{k_{u+2}},\dots,b_{k_r}$ from $\{b_1,b_2,\dots,b_m\}$.[2] Then b_i is said to be *indistinguishable* from b_j according to Θ. This may be expressed using an equivalence relation

$$b_i \sim_\Theta b_j$$

where the equivalence classes of $\sim_\Theta$ partition $b_1,\dots,b_m$ so that those in the same class are all indistinguishable from each other, but distinguishable from any member of another class, according to Θ. The *multiset* of the sizes of these equivalence classes is called the *spectrum* of Θ, denoted $\mathcal{S}(\Theta)$. The size r of this multiset will be called the *spectrum length* and denoted $|\mathcal{S}(\Theta)|$. The spectrum length of Θ is therefore the number of equivalence classes of $\sim_\Theta$. The spectrum consisting of m ones (corresponding to a state description where each of m constants is distinguishable from every other) will be denoted $\mathbf{1}_m$. The symbol $\emptyset$ will be used to denote the spectrum of a state description for zero constants (i.e. a tautology, by the above convention).

Given spectra $\tilde{m},\tilde{n}$ and a state description Θ with spectrum $\tilde{m}$ it can be shown, see [6], [9], [10], [15], that the number of state descriptions with spectrum $\tilde{n}$ extending

[2] Equivalently $\Theta \wedge b_i = b_j$ would be consistent were we to add equality to the language.

Θ depends only on $\tilde{m}, \tilde{n}$ and not on the particular choice of Θ. We denote this number by $\mathcal{N}(\tilde{m}, \tilde{n})$.

A principle of rationality, originally introduced in [10] and based on the idea that beyond their spectra, differences in state descriptions are irrelevant is that of

Spectrum Exchangeability, Sx

A probability function w on SL satisfies Spectrum Exchangeability if, for any state descriptions $\Theta(b_1, b_2, \ldots, b_m), \Phi(b_1, b_2, \ldots, b_m)$ such that $\mathcal{S}(\Theta) = \mathcal{S}(\Phi)$

$$w(\Theta) = w(\Phi).$$

This principle may be justified by the argument that, in the absence of any interpretation of the language, there is no reason to think any one state description more probable than any other of the same spectrum.[3]

The following principles are based on the identification of probability zero with the notion of impossibility.

Regularity, Reg

A probability function w on SL satisfies Regularity if $w(\phi) > 0$ for all consistent $\phi \in QFSL$.

Super-Regularity, SReg

A probability function w on SL satisfies Super-Regularity if $w(\theta) > 0$ for all consistent $\theta \in SL$.

Super-Regularity may be thought desirable on the grounds that, if zero probability is identified with impossibility, then any sentence which is consistent, and therefore theoretically possible, should receive non-zero probability. Put another way, a probability function which satisfies SReg gives probability 1 only to tautologies, and hence any other principle $\mathcal{P}$ which this probability function satisfies must have the minimal theory, just the set of tautologies. Notice that the properties Reg and SReg have the practical advantage of ensuring that conditional probabilities are always uniquely defined over consistent sentences from $QFSL$ and SL respectively.

[3] If L is a purely unary language then Sx is equivalent to the principle of Atom Exchangeability, see [15, chapter 14], which is in essence Carnap's *attribute symmetry*, see [1, p77].

The Theory of Sx

Let L be the first order language with finitely many relation symbols (possibly polyadic) as described above. It is shown in [6], [8], [9], [10], [15] that any probability function on SL which satisfies Sx may be expressed as a convex sum of probability functions of two basic types: *heterogeneous* and *homogeneous* functions, defined as follows.

Homogeneity

A probability function w on SL is homogeneous (abbreviated hom) if it satisfies Sx and for each $t \in \mathbb{N}^+$

$$\lim_{n\to\infty} w\left(\bigvee_{|\mathcal{S}(\Phi(a_1,\ldots,a_n))|=t} \Phi(a_1,\ldots,a_n)\right) = 0. \tag{2}$$

The disjunction is taken over all state descriptions of L for constants $a_1,\ldots,a_n$ with spectrum length t.

We now describe a particular important family of probability functions, the $u^{\bar{p},L}$ for $\bar{p} \in \mathbb{B}$, where

$$\mathbb{B} = \{\langle p_0, p_1, p_2, \ldots\rangle \mid p_1 \geq p_2 \geq \ldots \geq 0,\ p_0 \geq 0\ \&\ \sum_i p_i = 1\}.$$

For a given state description $\Phi(a_{i_1},\ldots,a_{i_m})$ and a given vector $\vec{c} \in \mathbb{N}^m$, Φ is said to be *consistent with* $\vec{c}$ if for $1 \leq s,t \leq m$, $c_s = c_t \neq 0 \implies a_{i_s} \sim_\Phi a_{i_t}$. The set of all state descriptions for $\vec{a}$ which are consistent with $\vec{c}$ is denoted $\mathcal{C}(\vec{c},\vec{a})$. For $\bar{p} \in \mathbb{B}$, the probability function $u^{\bar{p},L}$ is defined on a state description $\Phi(a_{i_1},\ldots,a_{i_m})$ by

$$u^{\bar{p},L}(\Phi(a_{i_1},\ldots,a_{i_m})) = \sum_{\substack{\langle c_1,\ldots,c_m\rangle\in\mathbb{N}^m \\ \Phi\in\mathcal{C}(\vec{c},\vec{a})}} |\mathcal{C}(\vec{c},\vec{a})|^{-1} \prod_{i=1}^{m} p_{c_i}. \tag{3}$$

It can be shown, see [7], [8] or [15, chapter 29] for the details, that $u^{\bar{p},L}$ now extends to a probability function on SL which will furthermore be homogeneous when

$$\bar{p} \in \mathbb{B}_\infty = \{\langle p_0, p_1, p_2, \ldots\rangle \in \mathbb{B} \mid p_0 > 0 \text{ or } p_i > 0 \text{ for all } i > 0\}.$$

Heterogeneity

For $t \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$, a probability function w on SL is t-heterogeneous (abbreviated t-het) if it satisfies Sx and

$$\lim_{n \to \infty} w \left(\bigvee_{|\mathcal{S}(\Phi(a_1,\ldots,a_n))|=t} \Phi(a_1, \ldots, a_n) \right) = 1. \tag{4}$$

Again, the disjunction is taken over all state descriptions of L for constants $a_1, \ldots, a_n$ with spectrum length t.

As with the homogeneous case we now describe a particular important family of t-heterogeneous probability functions, the $v^{\bar{p},L}$ for $\bar{p} \in \mathbb{B}_t$, where

$$\mathbb{B}_t = \{\langle p_0, p_1, p_2, \ldots, \rangle \in \mathbb{B} \mid p_0 = 0 \ \& \ p_t > 0 = p_{t+1}\}.$$

For $t \in \mathbb{N}^+$, let $\mathbb{N}_t = \{1, 2, \ldots, t\}$ and let $\bar{p} \in \mathbb{B}_t$. We define the probability function $v^{\bar{p},L}$ on state descriptions $\Phi(a_{i_1}, \ldots, a_{i_m})$ in terms of vectors $\vec{c} \in (\mathbb{N}_t)^m$ and a function $\mathcal{G}(\vec{c}, \Phi)$. For a fixed $\vec{c}$, if Φ is not consistent with $\vec{c}$, i.e. if for some $1 \leq s, t \leq m$, $c_s = c_t$ but $a_{i_s} \not\sim_\Phi a_{i_t}$, then $\mathcal{G}(\vec{c}, \Phi)$ is zero. Otherwise let $c_{g_1}, \ldots, c_{g_r}$ be the first instance of each distinct entry in $\vec{c}$ (supposing that $\vec{c}$ contains exactly r distinct entries, i.e. $|\{c_i \mid c_i \in \vec{c}\}| = r$) and let $\Phi' = \Phi[a_{i_{g_1}}, \ldots, a_{i_{g_r}}]$. Then $\mathcal{G}(\vec{c}, \Phi)$ takes the value

$$\frac{\mathcal{N}(\mathcal{S}(\Phi'), \mathbf{1}_t)}{\mathcal{N}(\emptyset, \mathbf{1}_t)},$$

i.e. the number of $\mathbf{1}_t$ extensions of Φ' as a proportion of the total number of $\mathbf{1}_t$ state descriptions for L. The value of $v^{\bar{p},L}(\Phi(a_{i_1}, \ldots, a_{i_m}))$ is defined in these terms to be

$$v^{\bar{p},L}(\Phi(a_{i_1}, \ldots, a_{i_m})) = \sum_{\langle c_1,\ldots,c_m\rangle \in (\mathbb{N}_t)^m} \mathcal{G}(\vec{c}, \Phi) \prod_{i=1}^{m} p_{c_i}. \tag{5}$$

It can be shown, see [7], [8] or [15, chapter 30] for the details, that $v^{\bar{p},L}$ now extends to a t-heterogeneous probability function on SL.

Probability functions satisfying Sx are a mixture of heterogeneous and homogeneous probability functions as the following theorem, see [6], [10], or [15, Theorem 30.2], explains.

The Ladder Theorem 1. *Any probability function w satisfying Sx can be expressed in the form*

$$w = \eta_0 w^{[0]} + \sum_{t=1}^{\infty} \eta_t w^{[t]}$$

where the $\eta_i \geq 0, \sum_i \eta_i = 1$, $w^{[0]}$ is homogeneous and $w^{[t]}$ is t-heterogeneous for $t > 0$.

In turn $w^{[0]}$ can be expressed as a convex combination of the basic homogeneous probability functions $u^{\bar{p},L}$, $\bar{p} \in \mathbb{B}_\infty$, in the sense of the forthcoming Theorem 14, whilst for $t > 0$ the $w^{[t]}$ can be expressed as convex combinations of the basic t-heterogeneous probability functions $v^{\bar{p},L}$, $\bar{p} \in \mathbb{B}_t$, for $t > 0$, in the sense of the forthcoming Theorem 2.

Our plan now is to investigate the theory of Sx, that is the set $Th(Sx)$ of sentences $\theta \in SL$ for which $w(\theta) = 1$ for all probability functions w on L satisfying Sx, by the separate investigation of the theory of t-heterogeneity, $Th(t\text{-}het)$, and the theory of homogeneity, $Th(hom)$.

The Theory of t-heterogeneity

We begin by stating some established results concerning t-heterogeneous probability functions, which will be needed later. The following theorem is proved in [8], or see [15, chapter 31], and explains the sense in which t-heterogeneous probability functions are convex mixtures of the basic t-heterogeneous probability functions $v^{\bar{p},L}$.

Theorem 2. *Let w be a t-heterogeneous probability function on SL. Then there is a measure[4] μ on the Borel[5] subsets of $\mathbb{B}_t$ such that*

$$w = \int_{\mathbb{B}_t} v^{\bar{p},L} \, d\mu(\bar{p}).$$

Conversely given such a measure μ, w defined as above is a t-heterogeneous probability function on SL.

The next result follows from the above theorem and the definition of the $v^{\bar{p},L}$ functions given in (5), and the fact that there will be a $\vec{c} \in (\mathbb{N}_t)^m$ consistent with $\Phi(b_1, \ldots, b_m)$ just if $|\mathcal{S}(\Phi)| \leq t$:

Lemma 3. *Let w be a t-heterogeneous probability function on SL.*

- *$w(\Phi(b_1, \ldots, b_m)) > 0$ for any state description $\Phi(b_1, \ldots, b_m)$ with spectrum length $|\mathcal{S}(\Phi)| \leq t$.*

[4] All measures will be taken to be countably additive and normalized.

[5] In other words the closure under complement and countable unions of the open subsets of, in this case, $\mathbb{B}_t$. This is sufficient to ensure that the functions $\bar{p} \mapsto v^{\bar{p},L}(\theta)$ are indeed integrable with respect to μ for $\theta \in SL$.

- *$w(\Psi(b_1,\ldots,b_m)) = 0$ for any state description $\Psi(b_1,\ldots,b_m)$ with spectrum length $|\mathcal{S}(\Psi)| > t$.*

It follows immediately that heterogenous functions do not satisfy Reg.

We now need a small technical result which is proved in [13, Lemma 1]. Let $t \in \mathbb{N}^+$, let g be the largest arity of any relation symbol in L, and let k be the largest of $t+1$ and g. Then for any $m \geq k$ and any state description $\Phi(a_1,\ldots,a_m)$ with spectrum length $|\mathcal{S}(\Phi)| \geq k \geq t+1$, there exists some s with $k \leq s \leq k+g = max\{t+1+g, 2g\}$, and some distinct $1 \leq i_1,\ldots,i_s \leq m$ such that

$$|\mathcal{S}(\Phi[a_{i_1},\ldots,a_{i_s}])| = s.$$

The significance of this is that if we let $s(t) = max\{t+1+g, 2g\}$, then any state description of L with spectrum length greater than t must have a restriction to $s(t)$ or fewer constants, with spectrum length greater than t. Therefore, the following (finite) sentence ζ_t may be used to express the idea that some state formula of spectrum length t is instantiated, and that any tuple of constants *of any length* will have to have spectrum length at most t.

Let ζ_t be the sentence

$$\bigvee_{\substack{\Theta(z_1,\ldots,z_t) \\ \mathcal{S}(\Theta)=\mathbf{1}_t}} \left(\exists x_1,\ldots,x_t\, \Theta(x_1,\ldots,x_t) \wedge \forall y_1,\ldots,y_{s(t)} \bigvee_{\sigma:\{y_1,\ldots,y_{s(t)}\}\to\{z_1,\ldots,z_t\}} \Theta_\sigma(y_1,\ldots,y_{s(t)}) \right)$$

where the outermost disjunction is over all state formulae with spectrum $\mathbf{1}_t$, and for $\sigma : \{y_1,\ldots,y_{s(t)}\} \to \{z_1,\ldots,z_t\}$ with image $\{z_{i_1},\ldots,z_{i_m}\}$, $\Theta_\sigma(y_1,\ldots,y_{s(t)})$ is the unique (up to logical equivalence) state formula $\Psi(y_1,\ldots,y_{s(t)})$ such that $\Psi(\sigma(y_1),\ldots,\sigma(y_{s(t)})) \equiv \Theta[z_{i_1},\ldots,z_{i_m}]$.

In more detail, ζ_t firstly says that there are some $a_{i_1}, a_{i_2},\ldots,a_{i_t}$ satisfying some state formula Θ of spectrum length t, so the a_{i_j} are all distinguished from each other by Θ. Additionally, ζ_t then says that if we take any $b_1, b_2,\ldots,b_{s(t)}$ then (taken together) they all look like clones of certain of the $a_{i_1}, a_{i_2},\ldots,a_{i_t}$. As a consequence of the choice of $s(t)$ this actually forces that *any* number of $b_1, b_2,\ldots,b_m$ must look like clones of certain of the $a_{i_1}, a_{i_2},\ldots,a_{i_t}$, in other words the universe has just t distinguishable elements in it (for example these $a_{i_1}, a_{i_2},\ldots,a_{i_t}$) and all the other elements are just clones of these.

Lemma 4. *If w is a t-heterogeneous probability function, then*

$$w(\zeta_t) = 1.$$

Proof. Suppose w is a t-heterogeneous probability function on SL and let $\Phi(a_1,\ldots,a_n)$ be a state description with spectrum length t. Then, by restricting Φ to one representative, $a_{g_1},\ldots,a_{g_t}$, of each equivalence class of $\sim_\Phi$, the result is a state description $\Theta(a_{g_1},\ldots,a_{g_t})$ with spectrum $\mathbf{1}_t$.

Furthermore, since the members of each equivalence class of $\sim_\Phi$ are indistinguishable from each other according to Φ, there is a unique surjective map $\sigma : \{a_1,\ldots,a_n\} \to \{z_1,\ldots,z_t\}$ such that $\Phi(a_1,\ldots,a_n) \equiv \Theta_\sigma(a_1,\ldots,a_n)$. So

$$\bigvee_{|\mathcal{S}(\Phi(a_1,\ldots,a_n))|=t} \Phi(a_1,\ldots,a_n) \models$$
$$\bigvee_{\mathcal{S}(\Theta(z_1,\ldots,z_t))=\mathbf{1}_t} \left(\exists x_1,\ldots,x_t\, \Theta(x_1\ldots,x_t) \wedge \bigvee_{\sigma:\{a_1,\ldots,a_n\}\to\{z_1,\ldots,z_t\}} \Theta_\sigma(a_1,\ldots,a_n) \right).$$

In fact, this last point applies not just to $\langle a_1,\ldots,a_n\rangle$, but to any tuple of constants of any length taken from $\{a_1,\ldots,a_n\}$, regardless of ordering or repeats. So for any tuple $\langle a_{i_1},\ldots,a_{i_{s(t)}}\rangle \in \{a_1,\ldots,a_n\}^{s(t)}$, there is always some mapping σ (not necessarily surjective) such that

$$\Phi[a_{i_1},\ldots,a_{i_{s(t)}}] \equiv \Theta_\sigma(a_{i_1},\ldots,a_{i_{s(t)}}),$$

and so

$$\bigvee_{|\mathcal{S}(\Phi(a_1,\ldots,a_n))|=t} \Phi(a_1,\ldots,a_n) \models \bigvee_{\mathcal{S}(\Theta(z_1,\ldots,z_t))=\mathbf{1}_t} (\exists x_1,\ldots,x_t\, \Theta(x_1\ldots,x_t)$$
$$\wedge \bigwedge_{i_1,\ldots,i_{s(t)}\leq n} \bigvee_{\sigma:\{a_{i_1},\ldots,a_{i_{s(t)}}\}\to\{z_1,\ldots,z_t\}} \Theta_\sigma(a_{i_1},\ldots,a_{i_{s(t)}})),$$

which, by standard properties of probability functions, see for example [11, Proposition 2.1(c)], gives that

$$w\left(\bigvee_{|\mathcal{S}(\Phi(a_1,\ldots,a_n))|=t} \Phi(a_1,\ldots,a_n) \right) \leq w\left(\bigvee_{\mathcal{S}(\Theta(z_1,\ldots,z_t))=\mathbf{1}_t} (\exists x_1,\ldots,x_t\, \Theta(x_1\ldots,x_t) \right.$$
$$\left. \wedge \bigwedge_{i_1,\ldots,i_{s(t)}\leq n} \bigvee_{\sigma:\{a_{i_1},\ldots,a_{i_{s(t)}}\}\to\{z_1,\ldots,z_t\}} \Theta_\sigma(a_{i_1},\ldots,a_{i_{s(t)}})) \right).$$

Taking the limit as $n\to\infty$ now gives

$$1 \leq w(\zeta_t)$$

by (4), since w is t-heterogeneous, and hence $w(\zeta_t)=1$ since $w(\zeta_t)\in[0,1]$. $\square$

It follows immediately from this that for any sentence $\theta \in SL$ such that $\zeta_t \models \theta$, $w(\theta) = 1$ for any t-heterogeneous w. We proceed to show that the converse also holds, after introducing some notation.

Let T be the set of $\mathbf{1}_t$ state formulae of L, i.e. $T = \{\Theta(z_1 \ldots, z_t) \mid \mathcal{S}(\Theta) = \mathbf{1}_t\}$, and define an equivalence relation $\approx$ on T by

$$\Theta \approx \Phi \iff \Theta(z_1 \ldots, z_t) \equiv \Phi(z_{\tau(1)} \ldots, z_{\tau(t)})$$

for some permutation τ of $\mathbb{N}_t$. Let $T_1, \ldots, T_u$ denote the equivalence classes of $\approx$, and for $1 \leq j \leq u$ let $\Theta_j \in T_j$ be some representative of its equivalence class.

Let

$$\eta_t^j = \exists x_1, \ldots, x_t \, \Theta_j(x_1, \ldots, x_t)$$

and

$$\xi_t^j = \forall y_1, \ldots, y_{s(t)} \bigvee_{\sigma : \{y_1, \ldots, y_{s(t)}\} \to \{z_1, \ldots, z_t\}} (\Theta_j)_\sigma(y_1, \ldots y_{s(t)})$$

and let

$$\zeta_t^j = \eta_t^j \wedge \xi_t^j.$$

Since

$$\exists x_1, \ldots, x_t \, \Theta(x_1, \ldots, x_t) \wedge \forall y_1, \ldots, y_{s(t)} \bigvee_{\sigma : \{y_1, \ldots, y_{s(t)}\} \to \{z_1, \ldots, z_t\}} \Theta_\sigma(y_1, \ldots, y_{s(t)}),$$

$$\exists x_1, \ldots, x_t \, \Phi(x_1, \ldots, x_t) \wedge \forall y_1, \ldots, y_{s(t)} \bigvee_{\sigma : \{y_1, \ldots, y_{s(t)}\} \to \{z_1, \ldots, z_t\}} \Phi_\sigma(y_1, \ldots, y_{s(t)})$$

are logically equivalent whenever $\Theta, \Phi \in T$ and $\Theta \approx \Phi$ it follows that ζ_t is equivalent to the disjunction of the pairwise disjoint sentences ζ_t^j

$$\zeta_t \equiv \bigvee_{j=1}^{u} \zeta_t^j = \bigvee_{j=1}^{u} (\eta_t^j \wedge \xi_t^j). \tag{6}$$

Let $M^j \in \mathcal{T}L$ be a model of ζ_t^j. Then $M^j \models \exists x_1, \ldots, x_t \, \Theta_j(x_1, \ldots, x_t)$, so suppose that $M^j \models \Theta_j(a_{g_1}, \ldots, a_{g_t})$. Since

$$M^j \models \xi_t^j, \tag{7}$$

for any constant symbol a_i there exists a unique $\sigma(a_i) \in \{a_{g_1}, \ldots, a_{g_t}\}$ such that

$$M^j \models (\Theta_j)_\sigma(a_{g_1}, \ldots, a_{g_t}, a_i).$$

Note that $\sigma(a_k) = a_k$ for $a_k \in \{a_{g_1}, \ldots, a_{g_t}\}$, and so $\sigma^2 = \sigma$.

Furthermore, the $\sigma(a_i)$ and a_i are indistinguishable in M^j, in the sense that for any state formula $\Phi(x_1, \ldots, x_{v+1})$ and $a_{k_1}, \ldots, a_{k_v}$

$$M^j \models \Phi(a_i, a_{k_1}, \ldots, a_{k_v}) \iff M^j \models \Phi(\sigma(a_i), a_{k_1}, \ldots, a_{k_v}). \tag{8}$$

For suppose otherwise. Then there must exist some constants $a_{k_1}, \ldots, a_{k_v}$ such that, for the unique state description $\Phi^+(a_i, \sigma(a_i), a_{k_i}, \ldots, a_{k_v})$ such that $M^j \models \Phi^+$,

$$a_i \not\sim_{\Phi^+} \sigma(a_i).$$

This would mean that, for the unique state description $\Psi(a_{g_1}, \ldots, a_{g_t}, a_i, a_{k_1}, \ldots, a_{k_v})$ such that $M^j \models \Psi$, $|\mathcal{S}(\Psi)| > t$, since $a_{g_1}, \ldots, a_{g_t}$ are all distinguishable in Ψ, and a_i is distinguishable from each of them. This contradicts (7) by the above discussion regarding the choice of $s(t)$.

This discussion now yields:

Lemma 5. *For any $\psi(a_1, \ldots, a_n) \in SL$*

$$M^j \models \psi(a_1, \ldots, a_n) \iff M^j \models \psi(\sigma(a_1), \ldots, \sigma(a_n)).$$

Proof. Straightforward by induction on the length of ψ using (8). □

We now define a new structure A_j for L with universe $|A^j| = \{a_{g_1}, \ldots, a_{g_t}\}$ by taking the interpretation of a_i in A^j to be $\sigma(a_i)$ and the interpretation of the relation symbol R_k of L in A^j to be the interpretation of R_k in M^j restricted to $|A^j|$. So essentially A^j is M^j with all the indistinguishable elements of M^j lumped together. Given this the following lemma is hardly surprising.

Lemma 6. *For any $\psi(a_1, \ldots, a_n) \in SL$*

$$M^j \models \psi(a_1, \ldots, a_n) \iff A^j \models \psi(a_1, \ldots, a_n).$$

Proof. Straightforward by induction on the length of ψ using Lemma 5 . □

Since $|A^j|$ is finite with every element named by a constant, when referring to the truth of sentences in A^j we can replace existential and universal quantifiers by finite disjunctions and conjunctions respectively. This observation gives us:

Lemma 7. *For any $n \in \mathbb{N}$ and any $\theta(a_1, \ldots, a_n) \in SL$, there exists a quantifier-free formula $\theta'(x_1, \ldots, x_t, a_1, \ldots, a_n)$ such that*

$$M^j \models \theta'(a_{g_1}/x_1, \ldots, a_{g_t}/x_t, a_1, \ldots, a_n) \iff M^j \models \theta(a_1, \ldots, a_n).$$

Proof. Let $\theta(a_1, \ldots, a_n) \in SL$ for some $n \in \mathbb{N}$ (possibly zero). Assume, without loss of generality, that θ is in Prenex Normal Form, so

$$\theta(a_1, \ldots, a_n) =$$
$$Q_p\, z_{p,1}, \ldots, z_{p,n_p}\, Q_{p-1}\, z_{p-1,1}, \ldots, z_{p-1,n_{p-1}}, \ldots, Q_1\, z_{1,1}, \ldots, z_{1,n_1}\, \phi(z_{1,1}, \ldots, z_{p,n_p}, a_1, \ldots, a_n),$$

where each Q_i is either $\forall$ or $\exists$ and $\phi(\vec{z}, a_1, \ldots, a_n) \in QFFL$.

Now let θ' be θ with each occurrence of

$$\forall z_{k,1}, \ldots, z_{k,n_k} \qquad \text{replaced by} \qquad \bigwedge_{\langle z_{k,1},\ldots,z_{k,n_k}\rangle \in \{x_1,\ldots,x_t\}^{n_k}}$$

and each occurrence of

$$\exists z_{k,1}, \ldots, z_{k,n_k} \qquad \text{replaced by} \qquad \bigvee_{\langle z_{k,1},\ldots,z_{k,n_k}\rangle \in \{x_1,\ldots,x_t\}^{n_k}}$$

for $k = 1, \ldots, p$, so that $\theta'(x_1, \ldots, x_t, a_1, \ldots, a_n)$ is a quantifier free formula mentioning constants from $a_1, \ldots, a_n$ and free variables $x_1, \ldots, x_t$ (only).

Then for M^j, A^j and σ as above,

$$\begin{aligned} M^j &\models \theta'(a_{g_1}/x_1, \ldots, a_{g_t}/x_t, a_1, \ldots, a_n) \\ &\iff A^j \models \theta'(a_{g_1}/x_1, \ldots, a_{g_t}/x_t, a_1, \ldots, a_n) \\ &\iff A^j \models \theta(a_1, \ldots, a_n) \\ &\iff M^j \models \theta(a_1, \ldots, a_n) \end{aligned} \tag{9}$$

where the first and third implications hold by Lemma 6 and the second holds since $|A^j| = \{a_{g_1}, \ldots, a_{g_t}\}$. □

Lemma 8. *For $\theta(a_1, \ldots, a_n) \in SL$ and some fixed $\sigma : \{z_1, \ldots, z_t, a_1, \ldots, a_n\} \to \{z_1, \ldots, z_t\}$*

$$\zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \ldots, a_n)) \models \theta(a_1, \ldots, a_n)$$

or

$$\zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \ldots, a_n)) \models \neg\theta(a_1, \ldots, a_n).$$

Proof. The result is clear if the left hand side is inconsistent. Assume otherwise, so by the remark following (7), σ must be the identity on the z_i. Let $M^j \in \mathcal{T}L$ be a model of ζ_t^j such that

$$M^j \models \zeta_t^j \wedge \exists \vec{z}\,(\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n)).$$

Then from Lemma 7, since the representatives $a_{g_1}, \dots, a_{g_t}$ were arbitrary up to satisfying Θ_j in M^j, we have that for θ' as given there,

$$M^j \models \forall \vec{z}\,(\Theta_j(\vec{z}) \to (\theta(a_1, \dots, a_n) \leftrightarrow \theta'(\vec{z}, a_1, \dots, a_n)))$$

regardless of the particular map σ. By earlier remarks concerning the completeness of the structures in $\mathcal{T}L$ this gives

$$\zeta_t^j \models \forall \vec{z}\,(\Theta_j(\vec{z}) \to (\theta(a_1, \dots, a_n) \leftrightarrow \theta'(\vec{z}, a_1, \dots, a_n))). \tag{10}$$

Since θ' is quantifier free and

$$\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n)$$

decides all the $\pm R_k(u_1, \dots, u_{r_k})$ for $u_1, \dots, u_{r_k}$ from $a_1, \dots, a_n, z_1, \dots, z_t$ it also decides θ', so

$$\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n) \models \theta' \quad \text{or} \quad \Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n) \models \neg\theta',$$

which we shall abbreviate to

$$\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n) \models \theta' \text{ or } \neg\theta'.$$

Hence by (10),

$$\zeta_t^j \wedge \Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n) \models \theta \text{ or } \neg\theta.$$

But now since $\vec{z}$ does not appear on the right hand side we obtain, as required, that

$$\zeta_t^j \wedge \exists \vec{z}\,(\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n)) \models \theta(a_1, \dots, a_n) \quad \text{or} \quad \neg\theta(a_1, \dots, a_n).$$

$\square$

Corollary 9. *If w is a t-heterogeneous probability function then w satisfies the Finite Values Property, FVP, meaning that for each n*

$$\{w(\theta(a_1, \dots, a_n)) \,|\, \theta(a_1, \dots, a_n) \in SL\}$$

is finite.

Proof. By Lemma 8, (6) and Lemma 4

$$w\left(\bigvee_{\langle j,\sigma\rangle\in A_\theta\cup A_{\neg\theta}} \zeta_t^j \wedge \exists\vec{z}\,(\Theta_j(\vec{z})\wedge(\Theta_j)_\sigma(\vec{z},a_1,\dots,a_n))\right)=1$$

where $A_\theta, A_{\neg\theta}$ are respectively the sets of pairs $\langle j,\sigma\rangle$ such that

$$\zeta_t^j \wedge \exists\vec{z}\,(\Theta_j(\vec{z})\wedge(\Theta_j)_\sigma(\vec{z},a_1,\dots,a_n)) \models \theta(a_1,\dots,a_n),$$

$$\zeta_t^j \wedge \exists\vec{z}\,(\Theta_j(\vec{z})\wedge(\Theta_j)_\sigma(\vec{z},a_1,\dots,a_n)) \models \neg\theta(a_1,\dots,a_n).$$

Hence

$$w\left(\bigvee_{\langle j,\sigma\rangle\in A_\theta} \zeta_t^j \wedge \exists\vec{z}\,(\Theta_j(\vec{z})\wedge(\Theta_j)_\sigma(\vec{z},a_1,\dots,a_n))\right)=w(\theta(a_1,\dots,a_n)),$$

leading to the conclusion that there can only be finitely many values of $w(\theta(a_1\dots,a_n))$, as $\theta(a_1,\dots,a_n)$ ranges over the sentences in SL mentioning only the constants $a_1,\dots,a_n$, since there are only finitely many possibilities for j and σ (for a fixed n) and hence only finitely many possible sets A_θ.

□

At the present time the status of the FVP, both from the standpoint of its mathematical consequences and its arguable rationality within the context of PIL, awaits further clarification. It seems at first sight rather demanding since although the number of constants in θ may be bounded, no such restriction is placed on the quantifier complexity of θ. On the other hand we already know of a number of examples where it does hold. In particular in the case of a purely unary language every probability function satisfies the FVP (for example by the identity (10.4) of [15]). For polyadic languages however it can fail (as we shall shortly see), indeed the required finiteness condition can hold for $a_1,\dots,a_n$ but fail for $a_1,\dots,a_n,a_{n+1}$. For more details see [5].

As far as the rationality of the FVP is concerned it is certainly to be commended as a *simplicity* requirement, from which an appeal to Occam's Razor may grant it some measure of rationality.[6]

We are now ready to prove the converse to our earlier observation that for w a t-heterogeneous probability function, if $\zeta_t^j \models \theta(\vec{a})$ then $w(\theta(\vec{a}))=1$.

[6]In fact, as explained in [5], this simplicity extends to a much deeper level than just the apparently contingent fact of only taking finitely many values.

Lemma 10. *If w is a t-heterogeneous probability function, $\theta(a_1, \dots, a_n) \in SL$ and $w(\theta(\vec{a})) = 1$ then $\zeta_t \models \theta(\vec{a})$.*

Proof. If $\zeta_t^j \not\models \theta(\vec{a})$ for some $1 \le j \le u$ then, since

$$\zeta_t^j \equiv \zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge \bigvee_\sigma (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n)),$$

there must be some σ for which

$$\zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n)) \not\models \theta(\vec{a}).$$

Hence this left hand side must be consistent and by Lemma 8

$$\zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n)) \models \neg\theta(\vec{a}). \tag{11}$$

By Lemma 3, $w(\zeta_t^j) > 0$ for each $1 \le j \le u$. So if

$$w(\zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n))) = 0$$

then it must be that

$$w((\Theta_j)_\sigma(a_1, \dots, a_n)) = 0.$$

But since $(\Theta_j)_\sigma(a_1, \dots, a_n)$ is a state description for $a_1, \dots, a_n$ with spectrum length at most t, this is false by Lemma 3. Therefore by (11),

$$w(\neg\theta(\vec{a})) \ge w(\zeta_t^j \wedge \exists \vec{z}\, (\Theta_j(\vec{z}) \wedge (\Theta_j)_\sigma(\vec{z}, a_1, \dots, a_n))) > 0.$$

The result follows. □

Since ζ_t does not mention any constants we immediately obtain from this lemma that:

Corollary 11. *If w is a t-heterogeneous probability function, $\theta(\vec{a}) \in SL$ and $w(\theta(\vec{a})) = 1$ then $\zeta_t \models \forall \vec{x}\, \theta(\vec{x})$ and so by Lemma 4, $w(\forall \vec{x}\, \theta(\vec{x})) = 1$.*

From Lemmas 4 and 10 we now obtain:

Theorem 12.

$$Th(t\text{-}het) = \{\theta \in SL \,|\, \zeta_t \models \theta\}.$$

Let $Th(Fin)$ denote the Theory of Finite Structures, that is the set

$$\{\theta \in SL \mid M \models \theta \text{ for every finite structure } M \text{ for } L\}.$$

We now show that this set is equal to the intersection over t of the theories of t-heterogeneity.

Theorem 13.

$$Th(Fin) = \bigcap_{t \in \mathbb{N}^+} Th(t\text{-}het).$$

Proof. Suppose that $\theta(a_1, \ldots, a_n) \in Th(Fin)$. Let $M^j \in \mathcal{T}L$ be an arbitrary model of ζ_t^j, and let A^j be defined in terms of M^j as above. Then $|A^j|$ is finite, so $A^j \models \theta(a_1, \ldots, a_n)$. By Lemma 6 then $M^j \models \theta(a_1, \ldots, a_n)$, so since t, j and M^j are arbitrary, by earlier remarks $\zeta_t \models \theta(a_1, \ldots, a_n)$ for each $t \in \mathbb{N}^+$. Therefore by Theorem 12, $\theta(a_1, \ldots, a_n) \in \bigcap_{t \in \mathbb{N}^+} Th(t\text{-}het)$.

Conversely suppose that $\theta(a_1, \ldots, a_n) \in Th(t\text{-}het)$ for each $t \in \mathbb{N}^+$. Let M be a finite structure for L, say M has exactly t distinguishable elements. Then a moment's thought shows that $M \models \zeta_t$, so $M \models \theta(a_1, \ldots, a_n)$ giving $\theta(a_1, \ldots, a_n) \in Th(Fin)$. The result follows.

$\square$

Notice that by Trakhtenbrot's Theorem, [16], $Th(Fin)$ is complete Π_1^0 so it certainly cannot be recursively axiomatized.

The Theory of Homogeneity

It is apparent from the definition of homogeneity (2) that there can be no homogeneous function on a purely unary language, since the spectrum length of any state description of a unary language can never exceed 2^q (where q is the number of predicate symbols in L). Therefore, assume for this section that the language L contains at least one non-unary relation symbol.

We begin by stating an established result concerning homogeneous probability functions, which will be used in this section. The following theorem, stating that any homogeneous probability function is a convex mixture of the basic homogeneous probability functions $u^{\bar{p},L}$ for $\bar{p} \in \mathbb{B}_\infty = \{\langle p_0, p_1, p_2, \ldots \rangle \in \mathbb{B} \mid p_0 > 0 \text{ or } p_i > 0 \text{ for all } i > 0\}$, and conversely, is proved in [8], or see [15, chapter 31].

Theorem 14. *Let w be a homogeneous probability function on SL. Then there is a measure μ on the Borel subsets of $\mathbb{B}_\infty$ such that*

$$w = \int_{\mathbb{B}_\infty} u^{\bar{p},L} \, d\mu(\bar{p}).$$

Conversely given such a measure μ, w defined as above is a homogeneous probability function on SL.

It is clear from (3) and the above theorem that any homogeneous probability function satisfies Reg, although we shall see later that it does not satisfy SReg.

We shall now show that $Th(hom)$, that is

$$\{\theta \in SL \mid w(\theta) = 1 \text{ for all homogeneous } w\},$$

is actually a variation on a well known complete theory, namely the theory of the random structure for L. In this sense then, our results forge a potentially fruitful connection between PIL and random structure theory. We will then refine this result to sentences mentioning just a fixed finite set of constants in order to show that homogeneous probability functions also satisfy the FVP.

Recall that r_e denotes the arity of relation R_e and let $\rho(z_1, z_2)$ be the formula

$$\bigwedge_{e=1}^{q} \bigwedge_{f=1}^{r_e} \forall x_1, \ldots, x_{f-1}, x_{f+1}, \ldots, x_{r_e}$$
$$(R_e(x_1, \ldots, x_{f-1}, z_1, x_{f+1}, \ldots, x_{r_e}) \leftrightarrow R_e(x_1, \ldots, x_{f-1}, z_2, x_{f+1}, \ldots, x_{r_e}))$$

which expresses that z_1 and z_2 are permanently indistinguishable from each other.

For $\vec{S} = S_1, \ldots, S_h$ a partition of $\mathbb{N}_m$, let $\upsilon^{\vec{S}}(y_1, \ldots, y_m)$ be

$$\bigwedge_{g=1}^{h} \bigwedge_{i,j \in S_g} \left(\rho(y_i, y_j) \wedge \bigwedge_{u \in \mathbb{N}_m - S_g} \neg\rho(y_i, y_u) \right), \tag{12}$$

and for $\Theta(x_1, \ldots, x_m)$ a state formula let

$$\Theta^{\vec{S}}(x_1, \ldots, x_m) = (\Theta(x_1, \ldots, x_m) \wedge \upsilon^{\vec{S}}(x_1, \ldots, x_m)).$$

Lemma 15. *Let w be a homogeneous probability function on language L. Then for a partition $\vec{S} = S_1, \ldots, S_h$ of $\mathbb{N}_m$, and $\Theta(x_1, \ldots, x_m)$, $\Psi(x_1, \ldots, x_m, x_{m+1})$ state formulae such that $\Psi \models \Theta$ and $\Psi^{\vec{S}, \{m+1\}}(x_1, \ldots, x_{m+1})$ consistent,*

$$w(\forall x_1, \ldots, x_m \, (\Theta^{\vec{S}}(x_1, \ldots, x_m) \to \exists x_{m+1} \, \Psi^{\vec{S}, \{m+1\}}(x_1, \ldots, x_m, x_{m+1}))) = 1. \tag{13}$$

Proof. By Theorem 14 it is sufficient to show the result for $w = u^{\bar{p}, L}$ where $\bar{p} \in \mathbb{B}_\infty$. Given $\bar{p} \in \mathbb{B}_\infty$, consider the following process for constructing a sequence of pairs, each consisting of a state description $\Phi_k(a_1, \ldots, a_k)$ and a sequence of 'colours'

$\vec{c}_k \in \mathbb{N}^k$. At stage $k = 0$ choose $\vec{c}_0 = \emptyset$, the empty sequence and $\Phi_0 = \top$, a tautology. At stage $k+1$ pick c_{k+1} from $\mathbb{N}$ with probability $p_{c_{k+1}}$, and then pick Φ_{k+1} from among those state descriptions consistent with $\vec{c}_{k+1}$ (i.e. those in $\mathcal{C}(\vec{c}_{k+1}, \langle a_1, \ldots, a_k, a_{k+1}\rangle))$ which extend Φ_k, according to the uniform distribution, i.e. with probability $|\mathcal{C}(\vec{c}_k, \langle a_1, \ldots, a_k\rangle)||\mathcal{C}(\vec{c}_{k+1}, \langle a_1, \ldots, a_k, a_{k+1}\rangle)|^{-1}$. (Note that, where $c_{k+1} = 0$ or is previously unseen in $\vec{c}_k$ there is a free choice of all those extensions Φ_{k+1} of Φ_k consistent with $\vec{c}_k$, while if $c_{k+1} > 0$ has occurred previously in $\vec{c}_k$, so that $c_{k+1} = c_r$, say, then Φ_{k+1} must be the unique extension of Φ_k such that a_{k+1} is a clone of a_r, meaning that $a_{k+1} \sim_{\Phi_{k+1}} a_r$.)

It is straightforward to show (as for example in [15, chapter 30]) that the probability that this process results at stage n in a particular pair $\langle \vec{c}_n, \Phi_n\rangle$ is given by

$$|\mathcal{C}(\vec{c}_n, \vec{a})|^{-1} \prod_{i=1}^{n} p_{c_i}.$$

Therefore, the value of

$$u^{\bar{p},L}(\Phi_n(a_1, \ldots, a_n)) = \sum_{\substack{\langle c_1, \ldots, c_n\rangle \in \mathbb{N}^n \\ \Phi_n \in \mathcal{C}(\vec{c}, \vec{a})}} |\mathcal{C}(\vec{c}, \vec{a})|^{-1} \prod_{i=1}^{n} p_{c_i}$$

is the sum, over all $\vec{c} \in \mathbb{N}^n$ consistent with Φ_n, of the probability of obtaining the pair $\vec{c}, \Phi_n$ by the process described.

Now suppose that this process has produced the pair $\langle \vec{c}, \Theta(a_1, \ldots, a_m)\rangle$ with probability

$$|\mathcal{C}(\vec{c}, \vec{a})|^{-1} \prod_{i=1}^{m} p_{c_i} \tag{14}$$

and $\vec{S}, \Psi$ are as in the statement of the lemma.

If it is the case that $c_i = c_j \neq 0$ just if i, j are in the same S_g, then for any new (previously unseen in $\vec{c}$) colour c_{m+1}, or 0, there is a fixed probability $1/C$ of picking $\Psi(a_1, \ldots, a_m, a_{m+1})$ as the next state description, where $C \geq SD(1, 2) > 1$ is the number of extensions allowed by the process. Similarly there is this same probability at each further choice where c_{m+s} is new, or 0, that the chosen state description will imply $\Psi(a_1, \ldots, a_m, a_{m+s})$. Since $p_0 > 0$ or there are infinitely many non-zero p_n, such a sequence of choices will, with probability 1, eventually produce a witness to $\exists x_{m+1}\, \Psi(a_1, \ldots, a_m, x_{m+1})$ which is assigned a different colour from those occurring in $\vec{c}$. Furthermore, with probability 1, any pair of constants assigned different colours become distinguishable from each other eventually.

Hence this probability (14) will all contribute to

$$u^{\bar{p},L}(\exists x_{m+1}\,(\Psi(a_1,\ldots,a_m,x_{m+1})\wedge\bigwedge_{j=1}^{m}\neg\rho(a_j,x_{m+1}))).$$

On the other hand, if for some i,j in different S_g we have $c_i=c_j\neq 0$ then no extension of $\langle\vec{c},\Theta(a_1,\ldots,a_m)\rangle$ can ever witness (12), while if for some i,j in the same S_g we have $c_i\neq c_j$, by this process a_i and a_j must eventually become distinguishable. In either case, the probability (14) will all contribute to

$$u^{\bar{p},L}(\neg\upsilon^{\vec{S}}(a_1,\ldots,a_m)).$$

Combining all of these probabilities now gives

$$u^{\bar{p},L}(\Theta^{\vec{S}}(a_1,\ldots,a_m)\rightarrow\exists x_{m+1}\,\Psi^{\vec{S},\{m+1\}}(a_1,\ldots,a_m,x_{m+1}))=1 \tag{15}$$

and by Ex this also holds for any distinct $a_{i_1},\ldots,a_{i_m}$ in place of $a_1,\ldots,a_m$. Where $b_1,\ldots,b_m\in\{a_1,a_2,\ldots\}^m$ are not all distinct, we may consider the restriction of $\Theta(b_1,\ldots,b_m)$ to its distinct arguments $\Theta[b_{j_1},\ldots,b_{j_s}]$, and find that (15) holds also with $b_1,\ldots,b_m$ in place of $a_1,\ldots,a_m$. The result now follows.

$\square$

Let Δ be the set of all sentences of the form (13) with $\vec{S},\Theta$ etc. as in Lemma 15. It follows immediately from the previous result that if $\Delta\models\phi$ then $w(\phi)=1$ for any homogeneous w. We now prove the converse to this.

Lemma 16. *If $\phi\in SL$ is such that $w(\phi)=1$ for some homogeneous probability function w on SL then $\Delta\models\phi$.*

Proof. Suppose that w, a homogeneous probability function on SL, and $\phi\in SL$ are such that $w(\phi)=1$. Let $L^{(m)}$ be L but with constant symbols a_i only for $i\leq m$. By a simple adaption of a well known back-and-forth argument of Fagin [2] (or see [11, Theorem 11.10] for a proof in the setting of this paper), for any state description $\Theta(a_1,\ldots,a_m)$ and partition $\vec{S}$, $\Delta\cup\{\Theta^{\vec{S}}(a_1,\ldots,a_m)\}$ is complete for $L^{(m)}$, i.e. decides any sentence of $SL^{(m)}$.

Hence if for any Θ, $\vec{S}$,

$$\Delta,\Theta^{\vec{S}}(a_1,\ldots,a_m)\not\models\phi$$

then it must be the case that $\upsilon^{\vec{S}}$ is consistent with Θ and

$$\Delta, \Theta^{\vec{S}}(a_1, \dots, a_m) \models \neg\phi. \tag{16}$$

Let $\bar{p} \in \mathbb{B}_\infty$. It follows from the definition of the $u^{\bar{p},L}$ functions given in (3) that $u^{\bar{p},L}(\Theta^{\vec{S}}(a_1, \dots, a_m)) > 0$ whenever $\upsilon^{\vec{S}}$ is consistent with Θ. Therefore, from (16), $u^{\bar{p},L}(\neg\phi) > 0$, so $u^{\bar{p},L}(\phi) < 1$. Hence by the first part of Theorem 14, $w(\phi) < 1$, a contradiction.

From this it follows that for every $\Theta, \vec{S}$

$$\Delta, \Theta^{\vec{S}}(a_1, \dots, a_m) \models \phi.$$

Hence

$$\Delta, \bigvee_{\Theta, \vec{S}} \{\Theta^{\vec{S}}(a_1, \dots, a_m)\} \models \phi,$$

giving

$$\Delta \models \phi(a_1, \dots, a_m), \tag{17}$$

since

$$\bigvee_{\Theta, \vec{S}} \Theta^{\vec{S}}(a_1, \dots, a_m)$$

is a tautology. □

From the previous two results we can now obtain the main theorem of this section:

Theorem 17.

$$Th(hom) = \{\theta \in SL \mid \Delta \models \theta\} = \{\theta \in SL \mid w(\theta) = 1 \textit{ for some homogeneous } w\}.$$

Corollary 18. *If w is a homogeneous probability function then w satisfies the Finite Values Property, FVP.*

Proof. The proof is essentially a repeat of that of Corollary 9, using the notation of Lemma 16. Suppose that w is a homogeneous probability function on SL. For any $\phi(a_1, \dots, a_m) \in SL$, since $\Delta \cup \{\Theta^{\vec{S}}(a_1, \dots, a_m)\}$ is complete for $L^{(m)}$, the consistent sentences $\Theta^{\vec{S}}$ may be partitioned into A_ϕ and $A_{\neg\phi}$, where A_ϕ is the set of $\Theta^{\vec{S}}(a_1, \dots, a_m)$ such that

$$\Delta, \Theta^{\vec{S}}(a_1, \dots, a_m) \models \phi(a_1, \dots, a_m)$$

(etc.). Since by Lemma 15, $w(\psi) = 1$ for $\psi \in \Delta$, and since

$$w\left(\bigvee_{\Theta^{\vec{S}}} \Theta^{\vec{S}}(a_1, \ldots, a_m)\right) = 1,$$

we must have

$$w(\phi) = w\left(\bigvee_{\Theta^{\vec{S}} \in A_\phi} \Theta^{\vec{S}}(a_1, \ldots, a_m)\right),$$

and the result follows as there are only finitely many possible A_ϕ as ϕ ranges over $SL^{(m)}$.

□

The Theory of Sx

By the Ladder Theorem 1 it is clear that the Theory of Sx must be equal to the intersection of $Th(hom)$ and $Th(t\text{-}het)$ for each $t \in \mathbb{N}^+$. By Theorem 13 this is equal to $Th(hom) \cap Th(Fin)$, and over the course of the next few lemmas we shall show that in fact this is just $Th(Fin)$.

Lemma 19. *When L is not purely unary,*

$$\lim_{t \to \infty} \frac{\mathcal{N}(\emptyset, \mathbf{1}_t)}{SD(t)} = 1.$$

Proof. Suppose a state description $\Phi(a_1, \ldots, a_t)$ is chosen at random from among all state descriptions for t constants. Then for any distinct $1 \leq i, j \leq t$, the probability that $a_i \sim_\Phi a_j$ is at most 2^{1-2t} (with equality when L consists of a single binary relation). The number of ways of choosing a distinct pair i, j is $t(t-1)/2$, so that the proportion of all state descriptions for $a_1, \ldots, a_t$ where some pair of constants is indistinguishable is bounded above by

$$\frac{t(t-1)}{2^{2t}}.$$

This value tends to zero as t tends to infinity, and the result follows.

□

The next two lemmas are aimed at showing that as $t \to \infty$ the $v^{\bar{p},L}$ for $\bar{p} \in \mathbb{B}_t$ get closer and closer to giving the sentences in Δ probability 1, and so also giving probability 1 to the (complete set of) logical consequences of Δ. This fact will then be used to derive the main theorem of this paper, Theorem 22.

Lemma 20. *For any $\delta > 0$ there exists $N \in \mathbb{N}$ such that for any $t \geq N$, any $\bar{p} \in \mathbb{B}_t$ and any $\theta \in SL$*

$$|u^{\bar{p},L}(\theta) - v^{\bar{p},L}(\theta)| < \delta.$$

Proof. Let $\delta > 0$ be fixed and let $\theta \in SL$. By a result of Landes [6, Theorem 12] (or see [15, Lemma 34.4]), for $t \in \mathbb{N}^+$ and any $\bar{p} \in \mathbb{B}_t$

$$u^{\bar{p},L}(\theta) = \frac{\mathcal{N}(\emptyset, \mathbf{1}_t)}{SD(t)} v^{\bar{p},L}(\theta) + \sum_G \frac{\mathcal{N}(\emptyset, 1_{|G|})}{SD(t)} v^{G(\bar{p}),L} \tag{18}$$

where $G = \{E_1, \ldots, E_{|G|}\}$ runs over the set of partitions of $\{1, \ldots, t\}$ with $|G| < t$ and $G(\bar{p}) \in \mathbb{B}_{|G|}$ has coordinates $\sum_{s \in E_i} p_s$. Since

$$\mathcal{N}(\emptyset, \mathbf{1}_t) + \sum_G \mathcal{N}(\emptyset, 1_{|G|}) = SD(t),$$

by Lemma 19, for t sufficiently large,

$$\frac{\mathcal{N}(\emptyset, \mathbf{1}_t)}{SD(t)} = 1 - \sum_G \frac{\mathcal{N}(\emptyset, 1_{|G|})}{SD(t)} > 1 - \delta$$

and the result follows by (18).

$\square$

Lemma 21. *Let $\vec{S}, \Theta, \Psi$ etc. be as in Lemma 15 and $\delta > 0$. Then for t sufficiently large and any $\bar{p} \in \mathbb{B}_t$*

$$v^{\bar{p},L}(\forall x_1, \ldots, x_m \, (\Theta^{\vec{S}}(x_1, \ldots, x_m) \to \exists x_{m+1} \, \Psi^{\vec{S},\{m+1\}}(x_1, \ldots, x_m, x_{m+1}))) > 1 - \delta.$$

Proof. By Lemma 20 it is enough to prove the result for $u^{\bar{p},L}$ in place of $v^{\bar{p},L}$.

The proof is a refinement of that for Lemma 15, the main difference being that where we had probability 1 in that lemma we will now only have probability close to 1. As a result we need to estimate just how close as we proceed.

Let t be fixed and large. As in the proof of Lemma 15 we can construct state descriptions $\Phi_k(a_1, \ldots, a_k)$ and sequences $\vec{c}_k$, this time from $(\mathbb{N}_t)^k$ since $\bar{p} \in \mathbb{B}_t$ is used, the probability that this process results at stage r in a particular pair $\vec{c}_r, \Phi_r$ being given by

$$|\mathcal{C}(\vec{c}_r, \vec{a})|^{-1} \prod_{i=1}^{r} p_{c_i}, \tag{19}$$

and

$$u^{\bar{p},L}(\Phi_r(a_1,\ldots,a_r)) = \sum_{\substack{\langle c_1,\ldots,c_r\rangle\in(\mathbb{N}_t)^r\\ \Phi_r\in\mathcal{C}(\vec{c},\vec{a})}} |\mathcal{C}(\vec{c},\vec{a})|^{-1}\prod_{i=1}^{r} p_{c_i}$$

is the sum, over each $\vec{c}\in(\mathbb{N}_t)^r$ consistent with Φ_r, of the probability of obtaining the pair $\vec{c},\Phi_r$ by the process described.

Let r be sufficiently large that

$$\sum_{\substack{\vec{c}\in(\mathbb{N}_t)^r\\ \{c_1,\ldots,c_r\}\subset\mathbb{N}_t}} \prod_{i=1}^{r} p_{c_i} = \sum_{j=1}^{t}(1-p_j)^r \le \frac{\delta}{3}. \tag{20}$$

holds.

Suppose that this process has produced the pair $\vec{c},\Phi_r$ where $\vec{c}=\langle c_1,\ldots,c_r\rangle\in(\mathbb{N}_t)^r$ and $\{c_1,\ldots,c_r\}=\mathbb{N}_t$, and $\Phi_r(a_1,\ldots,a_r)\in\mathcal{C}(\vec{c},\vec{a})$. Notice that since all the available colours $1,2,\ldots,t$ occur in $\vec{c}$, any continuation of this process can only produce clones of constants previously seen in Φ_r, so that Φ_{k+1} is uniquely determined by c_{k+1} and Φ_k for $k\ge r$.

Therefore, there is some state formula $\Upsilon(z_1,\ldots,z_t)$ and some distinct $g_1,\ldots,g_t\le r$ such that $c_{g_u}=u$ for $1\le u\le t$ and $\Phi_r\models\Upsilon(a_{g_1},\ldots,a_{g_t})$. Let χ be the sentence

$$\exists z_1,\ldots,z_t\,\Upsilon(z_1,\ldots,z_t)\wedge\forall y_1,\ldots,y_{s(t)} \bigvee_{\sigma:\{y_1,\ldots,y_{s(t)}\}\to\{z_1,\ldots,z_t\}} \Upsilon_\sigma(y_1,\ldots,y_{s(t)}) \tag{21}$$

where the notation is as defined in Section 3.1. Then any structure in $\mathcal{T}L$ which models this process for these particular $\vec{c}$ and Φ_r is a model of χ. (The only difference between χ and the ζ_t^j considered in the previous section is that Υ here has spectrum length *at most* t, not necessarily equal to t). It can be shown, exactly similarly as in the discussion around Lemma 8, that the sentence

$$\chi\wedge\Phi_r(a_1,\ldots,a_r)$$

is complete for $SL^{(r)}$. So in the case that

$$\chi\wedge\Phi_r(a_1,\ldots,a_r)\models\phi(a_1,\ldots,a_r),$$

we say ϕ is *fixed* by Φ_r, and all of the probability (19) will contribute to $u^{\bar{p},L}(\phi)$.

We now consider cases. If $\Phi_r(a_1, \ldots, a_r) \nvDash \Theta(a_1, \ldots, a_m)$ or $c_i = c_j$ for some $1 \leq i, j \leq m$ in different S_g then $\neg\Theta^{\vec{S}}(a_1, \ldots, a_m)$ is fixed.

Otherwise, if $c_i \neq c_j$ for some $1 \leq i, j \leq m$ in the same S_g, then each time a new colour c_k, $m < k \leq r$, was chosen there was a probability at most C^{-1}, where $C = SD(1, 2) > 1$ depends only on L, that the choice of $\Phi_k(a_1, \ldots, a_k)$ would not witness the failure of $\rho(a_i, a_j)$ (i.e. would not make a_i and a_j distinguishable). Hence the probability that such $\vec{c}, \Phi_r$ would *not* fix $\neg\Theta^{\vec{S}}(a_1, \ldots, a_m)$ is at most

$$\binom{m}{2}\left(\frac{1}{C}\right)^{t-m}.$$

Otherwise, if $\Phi_r \models \Theta(a_1, \ldots, a_m)$ and $\vec{c}$ is consistent with $\vec{S}$, with every choice of a new colour c_k, $m < k \leq r$, there is a probability at least D^{-1}, where $D = SD(m, m+1) > 1$ depends only on m and L, that

$$\Phi_k(a_1, \ldots, a_k) \models \Psi(a_1, \ldots, a_m, a_k)$$

and a probability of at most C^{-1} that a_k is indistinguishable from a_i for $1 \leq i \leq m$. This gives a probability of at most

$$\left(\frac{D-1}{D}\right)^{t-m} + m\left(\frac{1}{C}\right)^{t-m}$$

of $\vec{c}, \Phi_r$ *not* fixing $\exists x_{m+1}\, \Psi^{\vec{S},\{m+1\}}(a_1, \ldots, a_m, x_{m+1})$.

These estimates have been obtained for the specific constants $a_1, \ldots, a_m$. However, as discussed above, according to Φ_r there are at most t distinguishable constants among $a_1, \ldots, a_r$, and for each of the t^m choices of these (including those with repeated arguments) we have the same estimates as obtained for $a_1, \ldots, a_m$. Altogether then, the probability that Φ_r does not fix

$$\forall x_1, \ldots, x_m\, (\Theta^{\vec{S}}(x_1, \ldots, x_m) \to \exists x_{m+1}\, \Psi^{\vec{S},\{m+1\}}(x_1, \ldots, x_m, x_{m+1}))$$

is at most

$$t^m\left(\binom{m}{2}\left(\frac{1}{C}\right)^{t-m} + \left(\frac{D-1}{D}\right)^{t-m} + m\left(\frac{1}{C}\right)^{t-m}\right). \tag{22}$$

Hence to within the $\delta/3$ from (20), this same upper bound holds for $u^{\bar{p},L}$ and the result follows since (22) tends to zero as $t \to \infty$.

□

Theorem 22.

$$Th(Sx) = Th(Fin).$$

Proof. Since we already know by earlier remarks and Theorem 13 that

$$Th(Sx) = Th(hom) \cap \bigcap_{t \in \mathbb{N}^+} Th(t\text{-}het) = Th(hom) \cap Th(Fin),$$

it is enough to show that $Th(Fin) \subseteq Th(hom)$. So suppose $\eta(a_{i_1}, \dots, a_{i_m}) \in Th(Fin)$. Then by Theorem 13, $w(\eta(a_{i_1}, \dots, a_{i_m})) = 1$ for every t-heterogeneous w for each $t \in \mathbb{N}^+$. Therefore by Corollary 11, $w(\forall x_1, \dots, x_m\, \eta(x_1, \dots, x_m)) = 1$ for each such w. Hence

$$\forall x_1, \dots, x_m\, \eta(x_1, \dots, x_m) \in Th(Fin)$$

and

$$\eta(a_{j_1}, \dots, a_{j_m}) \in Th(Fin) \tag{23}$$

for any $j_1, \dots, j_m$, not necessarily distinct.

Again Δ is complete for sentences which do not contain constants so either

$$\Delta \models \forall x_1, \dots, x_m\, \eta(x_1, \dots, x_m) \ \text{ or } \ \Delta \models \neg \forall x_1, \dots, x_m\, \eta(x_1, \dots, x_m).$$

Suppose it was the latter. Then by Compactness there would be a finite subset $\{\phi_1, \dots, \phi_r\}$ of Δ such that

$$\phi_1, \dots, \phi_r \models \neg \forall x_1, \dots, x_m\, \eta(x_1, \dots, x_m). \tag{24}$$

By Lemma 21 for large enough t and $\bar{p} \in \mathbb{B}_t$,

$$v^{\bar{p},L}(\phi_i) > 1 - (2r)^{-1}$$

for each $1 \leq i \leq r$, so

$$v^{\bar{p},L}\left(\bigwedge_{i=1}^r \phi_i\right) = 1 - v^{\bar{p},L}\left(\bigvee_{i=1}^r \neg\phi_i\right) \geq 1 - \sum_{i=1}^r v^{\bar{p},L}(\neg\phi_i) > 1 - \sum_{i=1}^r (2r)^{-1} = 1/2.$$

From (24) then

$$v^{\bar{p},L}(\neg \forall x_1, \dots, x_m\, \eta(x_1, \dots, x_m)) > 1/2$$

so

$$v^{\bar{p},L}(\neg \eta(a_{j_1}, \dots, a_{j_m})) > 0$$

for some $a_{j_1}, \dots, a_{j_m}$, which contradicts (23) and Theorem 13.

Hence it must be that

$$\Delta \models \forall x_1, \dots, x_m \, \eta(x_1, \dots, x_m),$$

so

$$\Delta \models \eta(a_{i_1}, \dots, a_{i_m})$$

and $\eta(a_{i_1}, \dots, a_{i_m}) \in Th(hom)$ by Theorem 17, as required.

□

In fact $Th(hom)$ is a strict superset of $Th(Fin)$ since, for example, the sentence $\forall x(\Theta^{\{1\}}(x) \to \exists y \, \Psi^{\{1\},\{2\}}(x, y))$, where Ψ is an extension of Θ with spectrum length 2, is in $Th(hom)$ but is given probability 0 by any 1-heterogeneous probability function.

It is interesting to note that $Th(Sx)$ contains more than just tautologies. For example where L contains a binary relation (a similar example can be constructed for any polyadic relation), the conjunction ϕ of

$$\forall x_1 \neg R(x_1, x_1), \quad \forall x_1 \exists x_2 R(x_1, x_2),$$

$$\forall x_1, x_2, x_3((R(x_1, x_2) \wedge R(x_2, x_3)) \to R(x_1, x_3))$$

expresses that R is a strict partial ordering of the universe with no top element and therefore no finite model. Then $\neg\phi \in Th(Sx)$ and $w(\phi) = 0$ for any w satisfying Sx. This gives the following

Corollary 23. *For w a probability function on a not purely unary language L, if w satisfies Sx then w does not satisfy Super-Regularity.*

Concerning the status of the FVP and Sx, let w satisfy Sx and have the Ladder Theorem 1 representation

$$w = \eta_0 w^{[0]} + \sum_{t=1}^{\infty} \eta_t w^{[t]},$$

where the $\eta_0, \eta_t \geq 0$, $\eta_0 + \sum_{t=1}^{\infty} \eta_t = 1$, $w^{[0]}$ is homogeneous and for $t \geq 1$ $w^{[t]}$ is t-heterogeneous. Then by Corollaries 9 and 18, w will satisfy FVP if only finitely many of the η_t are non-zero. On the other hand, if infinitely many of the η_t are non-zero then, since their sum is bounded by 1, they must take infinitely many distinct values. Therefore, since $w(\zeta_t) = \eta_t$ for each $t \in \mathbb{N}^+$, the set $\{w(\zeta_t) \mid t \in \mathbb{N}^+\}$ is infinite, and w does not satisfy FVP.

Acknowledgement

We would like to thank the referees for their many useful suggestions and the typesetters for their forbearance.

References

[1] 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.

[2] Fagin, R., Probabilities on finite models, *Journal of Symbolic Logic*, 1976, **41**(1):50–58.

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

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

[5] Howarth, E., forthcoming Ph.D. Thesis, University of Manchester. To appear at http://www.maths.manchester.ac.uk/∼jeff/

[6] Landes, J., *The Principle of Spectrum Exchangeability within Inductive Logic*, Ph.D. Thesis, University of Manchester, 2009. Available at http://www.maths.manchester.ac.uk/∼jeff/

[7] Landes, J., Paris, J.B. & Vencovská, A., Some Aspects of Polyadic Inductive Logic, *Studia Logica*, 2008, **90**:3–16.

[8] Landes, J., Paris, J.B. & Vencovská, A., Representation Theorems for Probability Functions satisfying Spectrum Exchangeability in Inductive Logic, *International Journal of Approximate Reasoning*, 2009, **51**(1):35–55.

[9] Nix, C.J., *Probabilistic Induction in the Predicate Calculus*, Ph.D. Thesis, University of Manchester, UK, 2005. Available at http://www.maths.manchester.ac.uk/∼jeff/

[10] Nix C.J. & Paris, J.B., A Note on Binary Inductive Logic, *Journal of Philosophical Logic*, 2007, **36**(6):735–771.

[11] Paris, J.B., *The Uncertain Reasoner's Companion*, Cambridge University Press, 1994.

[12] 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

[13] Paris, J.B. & Rad, S.R., A Note on the Least Informative Model of a Theory, in *Programs, Proofs, Processes, CiE 2010,* eds. F.Ferreira, B.Löwe, E.Mayordomo, & L.M.Gomes, Springer LNCS 6158, 2010, pp342–351.

[14] Paris, J.B. & Vencovská, A., From Unary to Binary Inductive Logic, in *Games, Norms and Reasons – Logic at the Crossroads*, Eds. J van Bentham, A.Gupta, E. Pacuit, Synthese Library 353, Springer, 2011, pp199–212.

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

[16] Trakhtenbrot, B., The Impossibility of an Algorithm for the Decidability Problem on Finite Classes, *Proceedings of the USSR Academy of Sciences* (in Russian), 1950, **70**(4):569–572. [English translation in *AMS Translations, Series 2*, **23**:1–6, 1963.]