SYMMETRY PRINCIPLES IN POLYADIC INDUCTIVE LOGIC

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2016

Tahel Ronel
School of Mathematics

Contents

Word count 38446

The University of Manchester

Tahel Ronel
Doctor of Philosophy
Symmetry Principles in Polyadic Inductive Logic
February 8, 2016

We investigate principles of rationality based on symmetry in Polyadic Pure Inductive Logic. The aim of Pure Inductive Logic (PIL) is to determine how to assign probabilities to sentences of a language being true in some structure on the basis of rational considerations. This thesis centres on principles arising from instances of symmetry for sentences of first-order polyadic languages.

We begin with the recently introduced Permutation Invariance Principle (PIP), and find that it is determined by a finite number of permutations on a finite set of formulae. We test the consistency of PIP with established principles of the subject and show, in particular, that it is consistent with Super Regularity. We then investigate the relationship between PIP and the two main polyadic principles thus far, Spectrum Exchangeability and Language Invariance, and discover there are close connections. In addition, we define the key notion of polyadic atoms as the building blocks of polyadic languages. We explore polyadic generalisations of the unary principle of Atom Exchangeability and prove that PIP is a natural extension of Atom Exchangeability to polyadic languages.

In the second half of the thesis we investigate polyadic approaches to the unary version of Constant Exchangeability as invariance under signatures. We first provide a theory built on polyadic atoms (for binary and then general languages). We introduce the notion of a signature for non-unary languages, and principles of invariance under signatures, independence, and instantial relevance for this context, as well as a binary representation theorem. We then develop a second approach to these concepts using elements as alternative building blocks for polyadic languages.

Finally, we introduce the concepts of homomorphisms and degenerate probability functions in Pure Inductive Logic. We examine which of the established principles of PIL are preserved by these notions, and present a method for reducing probability functions on general polyadic languages to functions on binary languages.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

I am immeasurably grateful to my supervisors, Dr Alena Vencovská and Professor Jeff Paris, for their guidance, encouragement and support throughout my PhD. The results in this thesis are the product of working closely with Alena, from whom I have learnt a great amount. I could not have wished for better supervisors.

I would like to thank my colleagues, officemates and friends, in particular Liz Howarth, Malte Kließ, Inga Schwabrow and Rosie Laking, for sharing this experience with me and providing light relief in the more stressful times.

To my brothers and sister, and especially to my Mum, for her unwavering belief in my abilities, and for never allowing me to doubt myself. My Dad is the reason behind this PhD. He opened my eyes to the world of mathematics and taught me how to think. He will always be my inspiration. I would not have begun this journey, let alone completed it, without my parents.

Finally to my husband Tom, my biggest fan, my rock, for everything.

This PhD was generously funded by an Engineering and Physical Sciences Research Council Doctoral Training Award, for which I am grateful.

For my Dad

Chapter 1

Introduction

"Symmetry is what we see at a glance; based on the fact that there is no reason for any difference..." - B. Pascal, Pensées.

1.1 General Introduction

Decisions we make in everyday life often involve a degree of uncertainty; *which route should I take to get to work fastest? should I walk or take the bus? should I bring a jacket with me?* When trying to answer such questions logically, or rationally, we (perhaps subconsciously) rely on our previous experiences and on information we acquire that might influence our decision.

Consider, for example, the following situation. We invite a guest out for a meal. As the hosts, our guest asks us to recommend which of options A, B and C they should order. Not having any knowledge of their likes and dislikes, or the restaurant's strengths and weaknesses, how do we decide which dish to recommend? In this case, we do not favour any option over another, and are equally likely to pick any of A, B and C; we view the situation as completely *symmetric*. Suppose we now learn that our guest is a fan of the main ingredient of dish A. This is likely to enhance our belief they should order dish A. We have acquired some *relevant* information and it has affected our decision. We are then told that the restaurant was redecorated last year. Most of us would find this new information *irrelevant* to our decision, it should not increase or decrease our probability of choosing any of A, B or C.

The aim of Inductive Logic, beginning with Keynes' [18] and developed in the context we work in by Johnson [17] and Carnap [2, 3, 7, 9], is to supply us with a way to answer such questions. More precisely, it is to enable an agent to assign degrees of belief in a rational way, with rationality traditionally based on considerations of *symmetry*, *relevance* and *irrelevance*[1]. Of these, rational principles based on symmetry have featured most prominently in the subject and will form the principal theme of this thesis. This is due to the natural appeal symmetry possesses - most of us would readily accept that degrees of belief assigned to symmetric situations should be equal - and moreover, due to the potential of expressing symmetry formally.

This thesis is set in *Pure Inductive Logic* (PIL)[2], where the framework contains no specific interpretation. Namely, we imagine an agent inhabiting some structure for a language L (with no interpretation) who has no prior knowledge of what is true in this structure. Our task is to provide a rational way for the agent to allocate degrees of belief to sentences of L being true in this structure. Thus we are looking for a *belief function* defined on sentences of this language that satisfies possible requirements of rationality, in the form of mathematical statements we ask our function to satisfy.

Our approach in this investigation is mathematical rather than philosophical[3] in nature. We propose possible principles and investigate how they relate to each other, their (mathematical) consequences, and which 'belief functions' satisfy them. We do not claim a rational agent must, or even should, adhere to them. We merely present them as principles the agent might wish their function to satisfy. Moreover, we will concentrate our efforts on non-unary languages. The unary case has been thoroughly researched, while the polyadic symmetry picture is much less resolved. With the exception of some earlier explorations such as [16, 19], research into polyadic symmetry in the context of Inductive Logic has thus far focused on the principle of Spectrum Exchangeability [20, 22, 25, 31, 32]. We will investigate new aspects of polyadic symmetry, and by doing so hope to offer a new perspective on the area.

[1] More recently, arguments based on *analogy* have appeared, see [14, 15], [36, Chapter 22].

[2] Carnap made the distinction between Pure and Applied Inductive Logic in [5].

[3] For some references of philosophical perspectives on the classical principles see [36, page 7].

1.2 Mathematical Setting

We introduce the context and framework for this thesis, and the basic definitions, results and notation that will be used from the outset. Rather than providing all the required background in this section, we have provided in the first section of each chapter the theory relevant to that chapter. This has the advantage of the pertinent material being fresher in the reader's mind, as well as allowing us to present our investigation without delay.

We work with a first order language L containing variables $x_1, x_2, x_3, \dots$, constants $a_1, a_2, a_3, \dots$, finitely many relation symbols $R_1, R_2, \dots, R_q$ of finite arities $r_1, r_2, \dots, r_q$ respectively, and no function symbols nor the equality symbol. $t_1, t_2, t_3, \dots$ will denote terms of the language. The constants a_i are intended to exhaust the universe, in the sense that every individual in our universe can be represented by a constant from the a_i. We will use $b_1, b_2, \dots$ to denote a distinct choice of constants from the a_i; $y_1, y_2, \dots$ and $z_1, z_2, \dots$ for distinct choices of variables from the x_i. We identify the language L with the set $\{R_1, R_2, \dots, R_q\}$.

We say that a language is *unary* if it contains only unary predicate symbols; it is *r-ary* if all its relation symbols are at most r-ary and at least one is r-ary. If $r = 2$, we say *binary* rather than 2-ary. To emphasise the unary context where appropriate, we use symbols $P_1, P_2, \dots, P_q$ for unary predicates and L_q for the language containing just these predicate symbols.

SL will denote the set of sentences of L, $QFSL$ the quantifier-free sentences, and $(QF)FL$ the (quantifier-free) formulae of L. We will use Greek letters such as θ, ϕ, ψ for formulae of L, and intend that $\theta(x_1, x_2, \dots, x_n)$ implies that all the variables appearing in θ are amongst $x_1, x_2, \dots, x_n$. A similar convention applies to sentences $\theta(a_1, a_2, \dots, a_n)$. To simplify notation, we shall identify formulae which are logically equivalent throughout, and we will often use '$=$' rather than '$\equiv$' between logically equivalent formulae.

Let $\mathcal{T}L$ denote the set of structures of L with universe $\{a_1, a_2, a_3, \dots\}$, where each constant symbol a_i of L is interpreted in $\mathcal{M} \in \mathcal{T}L$ as $a_i \in \mathcal{M}$. We assume the structure the agent inhabits is one of the structures $\mathcal{M} \in \mathcal{T}L$, but the agent has no knowledge of what is true in $\mathcal{M}$.

We identify degree of belief with subjective probability[4] and the agent's 'belief function' with a *probability function*:

Definition 1.1. A function $w : SL \to [0,1]$ is a *probability function* on SL if 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) $w(\exists x\, \psi(x)) = \lim_{n\to\infty} w(\bigvee_{i=1}^{n} \psi(a_i))$.

Probability functions have a number of desirable properties[5]. Note in particular that logically equivalent sentences are given the same value by a probability function, and that convex sums of probability functions are probability functions.

Any probability function w satisfying just (P1) and (P2) on the quantifier-free sentences of L has a unique extension to a probability function on SL [13], so in many situations it suffices to think of probability functions as defined on quantifier-free sentences only, and satisfying (P1) and (P2). This can be further reduced[6] to a special class of such sentences called *state descriptions*:

Definition 1.2. The *state descriptions* of L are sentences $\Theta(b_1, \dots, b_n)$ of the form

$$\bigwedge_{d=1}^{q} \bigwedge_{\langle i_1 \dots, i_{r_d}\rangle \in \{1,\dots,n\}^{r_d}} \pm R_d(b_{i_1} \dots, b_{i_{r_d}}) \tag{1.1}$$

where $\pm R_d(b_{i_1} \dots, b_{i_{r_d}})$ denotes one of $R_d(b_{i_1} \dots, b_{i_{r_d}})$, $\neg R_d(b_{i_1}, \dots, b_{i_{r_d}})$.

These sentences completely describe how the constants $b_1, \dots, b_n$ behave in relation to each other (and no other constants). For $\Theta(b_1, \dots, b_n)$ a state description, $\Theta(x_1, \dots, x_n)$ is called a *state formula*. We make a convention that the state description on zero constants is a tautology and we denote it by '$\top$'. We use the upper case Greek letters Θ, Φ, Ψ to denote state descriptions and state formulae.

[4]For a justification of this approach see for example [36, Chapter 5], based on work by de Finetti [11] and Ramsey [38].

[5]Details can be found in [36, Chapter 3].

[6]See [36, Chapter 7] for an explanation.

So we have that a probability function is determined by its values on state descriptions. Moreover[7], any probability function w defined on state descriptions $\Theta(a_1, a_2, \ldots, a_n)$, $n \in \mathbb{N}$ to satisfy

(P1′) $w(\Theta(a_1, a_2, \ldots, a_n)) \geq 0$,

(P2′) $w(\top) = 1$,

(P3′) $w(\Theta(a_1, a_2, \ldots, a_n)) = \sum_{\Phi(a_1, a_2, \ldots, a_{n+1}) \models \Theta(a_1, a_2, \ldots, a_n)} w(\Phi(a_1, a_2, \ldots, a_{n+1}))$

extends, by the Disjunctive Normal Form Theorem, to a probability function on $QFSL$, and hence (uniquely) to a probability function on SL.

Notice that a state formula of the *unary* language L_q on n variables $\bigwedge_{i=1}^n \bigwedge_{d=1}^q \pm P_d(x_i)$ has the form

$$\bigwedge_{i=1}^n \alpha_{h_i}(x_i)$$

where $h_i \in \{1, \ldots, 2^q\}$ and $\alpha_1(x), \ldots, \alpha_{2^q}(x)$ are the formulae of the form

$$\pm P_1(x) \wedge \pm P_2(x) \wedge \cdots \wedge \pm P_q(x).$$

The α_j are known as the *atoms* of L_q, and they form the basic building blocks of sentences of a unary language. As such, they have featured prominently in Unary Inductive Logic, and have been used to formulate and investigate basic principles of the subject. We shall return to this point later on, in particular in Chapters 4 and 5.

Finally, we define the notion of *conditional probability*. Given a probability function w, the *conditional probability function* of $\theta \in SL$ given $\phi \in SL$, for ϕ such that $w(\phi) \neq 0$, is defined as

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

We adopt the convention that expressions like $w(\theta \mid \phi) = a$ stand for $w(\theta \wedge \phi) = a\, w(\phi)$ even if $w(\phi) = 0$. We assume throughout this investigation that if the agent assigns the subjective probability $w(\theta)$ to $\theta \in SL$ holding in $\mathcal{M}$, then the conditional probability $w(\theta \mid \phi)$ is what the agent supposes they would amend their choice of probability function to were they to learn that ϕ held in $\mathcal{M}$.

[7] [36, Chapter 7].

Rational Principles

We present a number of basic, established principles of PIL that we will use from the onset. As the investigation develops, we will introduce additional principles.

One of the most widely accepted principles in the subject pertains to the symmetry between the constants, stating that a rational probability function should treat the individual constants a_i equally. It can be stated as follows:

Constant Exchangeability, Ex

Let $\theta(a_1, \ldots, a_n) \in SL$ and let $b_1, \ldots, b_n$ be any other choice of distinct constant symbols from amongst the a_i. Then

$$w(\theta(a_1, \ldots, a_n)) = w(\theta(b_1, \ldots, b_n)). \tag{1.2}$$

Ex is sometimes imposed at the start of investigations in Inductive Logic as the first requirement a rational probability function should obey. We do not assume it here but we will explain the role it has in what follows. We remark that if a probability function satisfies Ex on the state descriptions of a language L then its extension to SL will also satisfy Ex[8].

A second principle based on symmetry which we shall come across relates to the symmetry between the relation symbols of the language. With no further knowledge, we have no reason to differentiate between two relation symbols of the same arity.

Predicate Exchangeability, Px

If R_i and R_j are relation symbols of L with the same arity, then

$$w(\theta) = w(\theta')$$

where θ' is the result of simultaneously swapping every occurrence of R_i in θ by R_j and every occurrence of R_j by R_i.

The next principle suggests that there is a symmetry between a relation symbol and its negation[9].

[8]For details, see [36, Chapters 6 and 7].

[9]where we use that $\neg\neg R$ is logically equivalent to R.

Strong Negation, SN

For $\theta \in SL$

$$w(\theta) = w(\theta')$$

where θ' is the result of replacing each occurrence of R in θ by $\neg R$.

The following principle is again based on symmetry but applies only to unary languages L_q. It refers to the symmetry between atoms in the zero knowledge situation. Polyadic approaches to this principle will be investigated in some detail in Chapter 4.

Atom Exchangeability, Ax

For any permutation τ of $\{1, 2, \dots, 2^q\}$ and constants $b_1, \dots, b_n$,

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\right) = w\left(\bigwedge_{i=1}^{n} \alpha_{\tau(h_i)}(b_i)\right).$$

The final symmetry principle we introduce at this time applies only to non-unary languages and concerns the symmetry between the variables in a relation.

Variable Exchangeability, Vx

Let R be an r-ary relation symbol of L, σ a permutation of $\{1, 2, \dots, r\}$. Then

$$w(\theta) = w(\theta')$$

where θ' is the result of replacing each occurrence of $R(t_1, t_2, \dots, t_r)$ in θ, where $t_1, \dots, t_r$ are any terms, by $R(t_{\sigma(1)}, t_{\sigma(2)}, \dots, t_{\sigma(r)})$.

We now mention two principles motivated by the idea that we should not dismiss as impossible sentences which could theoretically hold in $\mathcal{M}$.

Regularity, Reg

For any consistent $\theta \in QFSL$,

$$w(\theta) > 0.$$

And the stronger notion of Super Regularity,

Super Regularity, SReg

For any consistent $\theta \in SL$,

$$w(\theta) > 0.$$

Notation

We end this chapter with some notation and conventions which we will henceforth use without further explanation.

$\mathsf{S_n}$ will denote the permutation group of $\{1, 2, \ldots, n\}$. We will use '$\twoheadrightarrow$' for surjective functions and $\mathrm{Id} : A \to A$ for the function that maps every element of A to itself.

Instead of $\Theta(b_1, \ldots, b_n)$, we will at times write $\Theta(z_1, \ldots, z_n)(b_1, \ldots, b_n)$ to denote the result of simultaneously replacing each occurrence of z_i in $\Theta(z_1, \ldots, z_n)$ by b_i, $i = 1, \ldots, n$, and similarly for other substitutions. When we wish to make the individual substitutions clearer, we may also write $\Theta(z_1, \ldots, z_n)(b_1/z_1, b_2/z_2, \ldots, b_n/z_n)$ for this substitution.

Definition 1.3. For a state description $\Theta(b_1, \ldots, b_n)$ as in (1.1) and distinct $k_1, \ldots, k_g$ from $\{1, \ldots, n\}$,

$$\Theta(b_1, \ldots, b_n)[b_{k_1}, \ldots, b_{k_g}]\,,$$

or simply $\Theta[b_{k_1}, \ldots, b_{k_g}]$ when the context is clear, denotes the restriction of $\Theta(b_1, \ldots, b_n)$ to $b_{k_1}, \ldots, b_{k_g}$. That is, the conjunction of the literals from (1.1) with $\{i_1, \ldots, i_{r_d}\} \subseteq \{k_1, \ldots, k_g\}$. We define the restriction of a state formula similarly.

Example. Let L contain a single binary relation symbol R and let $\Theta(b_1, b_2, b_3)$ be the conjunction of

$$\begin{array}{ccc} R(b_1, b_1) & \neg R(b_1, b_2) & R(b_1, b_3) \\ R(b_2, b_1) & R(b_2, b_2) & \neg R(b_2, b_3) \\ \neg R(b_3, b_1) & R(b_3, b_2) & \neg R(b_3, b_3)\,. \end{array}$$

Then $\Theta(b_1, b_2, b_3)[b_1, b_2]$ is the conjunction of

$$\begin{array}{cc} R(b_1, b_1) & \neg R(b_1, b_2) \\ R(b_2, b_1) & R(b_2, b_2)\,. \end{array}$$

Definition 1.4. For $\sigma : \{y_1, \ldots, y_n\} \twoheadrightarrow \{z_1, \ldots, z_m\}$ and a state formula $\Theta(z_1, \ldots, z_m)$, there is a unique state formula (up to logical equivalence) $\Phi(y_1, \ldots, y_n)$ such that

$$\Phi(\sigma(y_1), \ldots, \sigma(y_n)) \equiv \Theta(z_1, \ldots, z_m).$$

We denote this Φ by

$$(\Theta(z_1,\dots,z_m))_\sigma(y_1,\dots,y_n)$$

or more simply by Θ_σ if the variables are clear.

Let L be a language containing *a single binary relation symbol* R. Any state description of L on n constants $b_1,\dots,b_n$ may be represented by an $n\times n$ $\{0,1\}$-matrix where 1 or 0 at the (i,j)th entry means this state description implies $R(b_i,b_j)$ or $\neg R(b_i,b_j)$ respectively. State formulae of this language are represented similarly.

Example. The state formula $\Theta(z_1,z_2)$ *corresponding to the conjunction of*

$$\begin{array}{ll} R(z_1,z_1) & \neg R(z_1,z_2) \\ R(z_2,z_1) & \neg R(z_2,z_2) \end{array}$$

may be represented by the matrix

$$\begin{array}{cc} 1 & 0 \\ 1 & 0 \end{array}.$$

Let $\sigma:\{y_1,y_2,y_3\}\twoheadrightarrow\{z_1,z_2\}$ *be such that* $\sigma(y_1)=\sigma(y_3)=z_1$, $\sigma(y_2)=z_2$. *Then the state formula* Θ_σ *can be represented by the matrix*

$$\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{array}.$$

y_1 *and* y_3 *are both 'clones' of* z_1, y_2 *is a 'clone' of* z_2.

Finally, we mention two particular probability functions that will come up in the investigation that ensues. When applied to unary languages, they form the end points of Carnap's Continuum of Inductive Methods (see [4, 7, 9, 17], [36, Chapters 16, 24]).

- The probability function c^L_∞, also known as the *completely independent probability function*, treats each $\pm R(b_1,\dots,b_r)$, where r is the arity of R, as stochastically independent and occurring with probability $\frac{1}{2}$.

- The probability function c^L_0 believes all constants behave in the same way[10]. That is, for each of the 2^q possible assignments of $\pm$, it satisfies

$$c^L_0\left(\bigwedge_{d=1}^{q}\forall x_1,\dots,x_{r_d}\pm R_d(x_1,x_2,\dots,x_{r_d})\right)=2^{-q}.$$

[10] In unary, c^L_0 believes all constants will satisfy the same atom as the first one seen (and gives probability 2^{-q} to such state descriptions). In contrast, c^L_∞ involves no learning from experience.

Chapter 2

Properties of the Permutation Invariance Principle

2.1 Introduction

We begin this chapter by presenting the formal framework for polyadic *symmetry* in Pure Inductive Logic, as set out in [35] and [36, Chapter 39]. This will allow us to introduce the Permutation Invariance Principle, the main object of investigation for the first part of this thesis, which first appeared in [35]. We then proceed to explore some of the properties this principle possesses. Lemma 2.2 in many ways underpins much of the investigation that follows, and the results from Section 2.4 will be useful later on, in Chapter 4. Results from this chapter appear also in [39] and in [36, Chapter 41].

Let L be an r-ary language and let $\mathcal{T}L$ denote the set of structures for L as defined on page 11. Let BL be the two-sorted structure with universe $\mathcal{T}L$, the sets

$$[\theta] = \{\mathcal{M} \in \mathcal{T}L \mid \mathcal{M} \models \theta\}$$

for $\theta \in SL$ and the membership relation between elements of $\mathcal{T}L$ and these sets.

An *automorphism* η of BL is a bijection of $\mathcal{T}L$ such that for each $\theta \in SL$,

$$\eta[\theta] = \{\eta\mathcal{M} \mid \mathcal{M} \in \mathcal{T}L\,,\, \mathcal{M} \models \theta\} = [\phi]$$

for some $\phi \in SL$ and conversely, for each $\phi \in SL$,

$$\eta^{-1}[\phi] = \{\eta^{-1}\mathcal{M} \mid \mathcal{M} \in \mathcal{T}L\,,\, \mathcal{M} \models \phi\} = [\theta]$$

for some $\theta \in SL$.

We will henceforth write[1] $\eta(\theta)$, or just $\eta\theta$, for the sentence $\phi \in SL$ such that $\eta[\theta] = [\phi]$. Note that θ, ϕ are determined up to logical equivalence only, however this should not be a problem for us, since we are identifying logically equivalent sentences/formulae throughout this account.

As is customary in investigations of Pure Inductive Logic, we assume a rational agent is aware of the structure BL, inhabits one of the structures $\mathcal{M} \in \mathcal{T}L$ but is unaware of which particular $\mathcal{M}$ it is. When the agent chooses their rational probability function w, it would therefore be reasonable to assume that justification for the probability $w(\theta)$ for $\theta \in SL$ (equivalently $[\theta] \in BL$) should apply also to $w(\eta\theta)$ for any automorphism η of BL. In other words, we are identifying a 'symmetry' of L with an automorphism of BL. This gives us the following symmetry principle[2] for a probability function w on SL:

The Invariance Principle, INV

For any automorphism η of BL and $\theta \in SL$

$$w(\theta) = w(\eta\theta).$$

We remark that for any probability function w on sentences of L and an automorphism η of BL, $w_\eta : SL \to [0, 1]$ given by $w_\eta(\theta) = w(\eta\theta)$ is a probability function on SL. This can be seen by checking that conditions (P1)-(P3) from Page 12 hold for w_η, see[3] [36, Chapter 23].

INV is rather strong a principle and can be shown to contain as special cases many of the symmetry principles traditionally studied in the subject [34, 36]. In particular, INV encompasses the Principle of Constant Exchangeability, a point we shall expand on in Section 2.3. In fact, previous investigations into INV for probability functions on *unary* languages have proven INV to be *too strong* a principle, leaving only one[4] (somewhat

[1] thus avoiding overuse of square brackets, which also denote restrictions of formulae, see page 16. This notation is now established so we keep to it; it should be clear from the context what is meant.

[2] This principle for unary L first appeared in [34], and for polyadic L in [35].

[3] This reference addresses the case when L is just unary, however the proof works also for polyadic languages L.

[4] This is Carnap's c_0^L, as described on page 17. We will encounter the polyadic version of this probability function later on, on pages 80, 131.

unsatisfactory) function that satisfies it [34]. On the other hand, it is not yet clear what its full effect for general languages is. The reason for INV eliminating nearly all probability functions in the unary context is that some automorphisms can force state descriptions with different numbers of constants to have the same probabilities, which - combined with all other conditions INV imposes - is almost never satisfied.

This raises the question of what happens if we require our automorphisms to map state descriptions to state descriptions respecting the number of constants, and what the probability functions that satisfy this weaker version of INV would be, where we only demand that $w(\theta) = w(\eta\theta)$ for $\theta \in SL$ and for such automorphisms η. It turns out [35] that such automorphisms must be in a certain sense uniform and up to a permutation of all constants, they must be of the type described below.

We say that a function $\digamma$ *permutes state formulae* if for each n and (distinct) variables $z_1, \dots, z_n$, $\digamma$ permutes the state formulae $\Theta(z_1, \dots, z_n)$ in these variables.

Definition 2.1. An *automorphism η of BL permutes state formulae* if there is a function $\bar{\eta}$ that permutes state formulae such that for any $b_1, \dots, b_n$ and state formulae $\Theta(z_1, \dots, z_n)$

$$\eta(\Theta(b_1, \dots, b_n)) = \bar{\eta}(\Theta(z_1, \dots, z_n))(b_1, \dots, b_n),$$

where $\bar{\eta}(\Theta(z_1, \dots, z_n))(b_1, \dots, b_n)$ is the state description arrived at by applying $\bar{\eta}$ to $\Theta(z_1, \dots, z_n)$ and then substituting $b_1, \dots, b_n$ into the resulting state formula.

Let $\digamma$ be a function permuting state formulae and satisfying the following conditions from [35]:

(A) *For each state formula $\Theta(z_1, \dots, z_m)$ and mapping $\sigma : \{y_1, \dots, y_n\} \twoheadrightarrow \{z_1, \dots, z_m\}$,*

$$(\digamma(\Theta))_\sigma = \digamma(\Theta_\sigma),$$

where σ is surjective and Θ_σ is the unique state formula[5] *$\Psi(y_1, \dots, y_n)$ such that $\Psi(\sigma(y_1), \dots, \sigma(y_n)) = \Theta(z_1, \dots, z_m)$.*

(B) *For each state formula $\Theta(z_1, \dots, z_m)$ and distinct $i_1, \dots, i_k \in \{1, \dots, m\}$*

$$\digamma(\Theta)[z_{i_1}, \dots, z_{i_k}] = \digamma(\Theta[z_{i_1}, \dots, z_{i_k}]),$$

[5] defined on page 16.

where $\Theta[z_{i_1}, \ldots, z_{i_k}]$ *is the restriction*[6] *of* $\Theta(z_1, \ldots, z_m)$ *to these variables.*

Note that where no confusion may arise, we write Θ in place of $\Theta(z_1, \ldots, z_n)$ and $\digamma(\Theta)$ for $\digamma(\Theta(z_1, \ldots, z_n))$ in favour of clarity of notation.

By Theorems 1 and 2 of [35], every function $\digamma$ that permutes state formulae and satisfies conditions (A) and (B) is a function $\bar{\eta}$ for some automorphism η of BL that permutes state formulae, and conversely, for every automorphism η that permutes state formulae, $\bar{\eta}$ satisfies (A) and (B).

We are now finally in a position to formally state the Permutation Invariance Principle from [35]. Restricting the Invariance Principle to include only the automorphisms of BL that permute state formulae gives us

The Permutation Invariance Principle, PIP

For any permutation of state formulae $\digamma$ *that satisfies* (A) *and* (B) *and a state description* $\Theta(b_1, \ldots, b_n)$

$$w(\Theta(b_1, \ldots, b_n)) = w(\digamma(\Theta)(b_1, \ldots, b_n)).$$

2.2 A Finite Characterisation of PIP

The following lemma shows that the Permutation Invariance Principle can be equivalently stated to involve invariance under *finitely many permutations*, specified by their action on a particular *finite set of formulae.* This set of formulae will form a fundamental component of much of this thesis.

Lemma 2.2. *Let* $\digamma$ *be a function that permutes state formulae and satisfies* (A) *and* (B). *Then* $\digamma$ *is uniquely determined by its action on state formulae of* r *variables, where* r *is the highest arity of an* L*-relation symbol.*

Proof. Consider a state formula $\Psi(z_1, \ldots, z_s)$ where $s < r$ and let $\Theta(z_1, \ldots, z_r)$ be such that $\Theta \models \Psi$. By condition (B)

$$\digamma(\Psi) = \digamma(\Theta[z_1, \ldots, z_s]) = \digamma(\Theta)[z_1, \ldots, z_s]$$

[6]defined on page 16.

so the values of F on state formulae of fewer than r variables are determined by its values on state formulae of r variables.

Now let $\Psi(z_1, \ldots, z_r, \ldots, z_n)$ be a state formula with $n > r$ and suppose there is a function F_1 that permutes state formulae and satisfies (A) and (B), such that $F(\Theta) = F_1(\Theta)$ for all state formulae Θ on r variables, but $F(\Psi) \neq F_1(\Psi)$. Then there must be a relation symbol R_d of L and (not necessarily distinct) $z_{i_1}, \ldots, z_{i_{r_d}}$ from $\{z_1, \ldots, z_n\}$ such that

$$F(\Psi) \models R_d(z_{i_1}, \ldots, z_{i_{r_d}}) \quad \text{and} \quad F_1(\Psi) \models \neg R_d(z_{i_1}, \ldots, z_{i_{r_d}}) \tag{2.1}$$

or vice versa. Let $z_{k_1}, \ldots, z_{k_r}$ be distinct variables from $\{z_1, \ldots, z_n\}$ such that all of $z_{i_1}, \ldots, z_{i_{r_d}}$ are included amongst them.

By condition (B) and since F, F_1 agree on state formulae of r variables, we have

$$F(\Psi)[z_{k_1}, \ldots, z_{k_r}] = F(\Psi[z_{k_1}, \ldots, z_{k_r}]) = F_1(\Psi[z_{k_1}, \ldots, z_{k_r}]) = F_1(\Psi)[z_{k_1}, \ldots, z_{k_r}],$$

contradicting (2.1). Thus the claim holds also for state formulae with more than r variables, as required. □

As an immediate consequence of this lemma, the set

$$\mathcal{F} = \{F \mid F \text{ permutes state formulae and satisfies (A) and (B)}\}$$

is finite and in the next section we will show that we can therefore generate a probability function w' that satisfies PIP from an arbitrary probability function w by averaging over 'permuted versions' of w. Furthermore, Proposition 2.4 shall show that w' will preserve some characteristic properties of w and thus bear witness to their compatibility with PIP.

2.3 PIP and Other Principles

We use the result of the previous section to test the consistency of the Permutation Invariance Principle with some long standing principles of Pure Inductive Logic. In particular, we will focus on the rather elusive principle of Super Regularity and clarify its status with respect to both PIP and INV.

Before doing so however, we mention some of the relationships we are already aware of between INV and PIP and other rational principles of PIL. Firstly, the commonly assumed Principle of Constant Exchangeability is implied by INV but not by PIP. In order to see that INV implies Ex [35], let $\sigma \in \mathsf{S}_{\mathbb{N}^+}$ and let $\eta\mathcal{M}$ be formed from $\mathcal{M} \in \mathcal{T}L$ by replacing each a_i in $\mathcal{M}$ by $a_{\sigma(i)}$. Then η extends to an automorphism of BL and requiring $w(\theta) = w(\eta\theta)$ for this η gives Ex. On the other hand, it is not surprising that PIP does not imply Ex since PIP requires the probability of two state descriptions instantiating the *same constants* to get the same probability, and makes no reference to the same sentence acting on different constants.

We can show that Ex is not implied by PIP explicitly by providing a probability function that satisfies PIP but not Ex. For this purpose, we touch on a topic we will explore in detail in Chapter 4. For the time being however, it suffices to note that when L is purely unary, PIP is equivalent to the Principle of Atom Exchangeability (see page 15): in unary, the functions $\digamma$ permuting state formulae and satisfying (A) and (B) are exactly those generated by permutations of atoms. So we can justify our claim by suggesting a probability function on unary languages that satisfies Ax but not Ex.

Let L_q be a unary language containing q predicate symbols and let $w^{\delta_1}_{L_q}$, $w^{\delta_2}_{L_q}$ be functions from the NP-continuum[7], as described in [29, 30] and in Chapters 18 and 19 of [36], with δ_1, δ_2 not both 0. Define

$$v\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = w^{\delta_1}_{L_q}\left(\bigwedge_{\substack{1\le i\le m\\ i \text{ odd}}} \alpha_{h_i}(a_i)\right) \cdot w^{\delta_2}_{L_q}\left(\bigwedge_{\substack{1\le i\le m\\ i \text{ even}}} \alpha_{h_i}(a_i)\right).$$

Then v extends to a probability function on SL_q since it satisfies (P1′)-(P3′) from page 13: (P1′) and (P2′) clearly hold. To check that (P3′) holds, suppose that $m+1$ is even. The case when $m+1$ is odd follows similarly with the roles of δ_1 and δ_2 interchanged. Note that $\Phi(a_1, \ldots, a_m, a_{m+1}) \models \Theta(a_1, \ldots, a_m)$ just when

[7]These functions can be defined as follows. Fix δ such that $-(2^q-1)^{-1} \le \delta \le 1$ and let $\gamma = 2^{-q}(1-\delta)$. Then for a state description $\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)$ of L_q

$$w^{\delta}_{L}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = 2^{-q}\sum_{j=1}^{2^q} \gamma^{m-m_j}(\gamma+\delta)^{m_j}$$

where $m_j = |\{i \,|\, h_i = j\}|$.

$\Phi(a_1, \ldots, a_m, a_{m+1}) = \Theta(a_1, \ldots, a_m) \wedge \alpha_j(a_{m+1})$ for $j \in 1, \ldots, 2^q$. So we have

$$\begin{aligned}
\sum_{\Phi \models \Theta} v(\Phi) &= \sum_{j=1}^{2^q} v\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \wedge \alpha_j(a_{m+1})\right) \\
&= \sum_{j=1}^{2^q} w_{L_q}^{\delta_1}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ odd}}} \alpha_{h_i}(a_i)\right) w_{L_q}^{\delta_2}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ even}}} \alpha_{h_i}(a_i) \wedge \alpha_j(a_{m+1})\right) \\
&= w_{L_q}^{\delta_1}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ odd}}} \alpha_{h_i}(a_i)\right) \sum_{j=1}^{2^q} w_{L_q}^{\delta_2}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ even}}} \alpha_{h_i}(a_i) \wedge \alpha_j(a_{m+1})\right) \\
&= w_{L_q}^{\delta_1}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ odd}}} \alpha_{h_i}(a_i)\right) w_{L_q}^{\delta_2}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ even}}} \alpha_{h_i}(a_i)\right) \\
&= v(\Theta),
\end{aligned}$$

where the penultimate equality follows since $w_{L_q}^{\delta_2}$ satisfies (P3′) and going through all the j's gives all the state descriptions that extend the state description $\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ even}}} \alpha_{h_i}(a_i)$.

Furthermore, v satisfies Ax but not Ex. Let $\sigma \in \mathsf{S}_{2^q}$.

$$\begin{aligned}
v\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) &= w_{L_q}^{\delta_1}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ odd}}} \alpha_{h_i}(a_i)\right) \cdot w_{L_q}^{\delta_2}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ even}}} \alpha_{h_i}(a_i)\right) \\
&= w_{L_q}^{\delta_1}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ odd}}} \alpha_{\sigma(h_i)}(a_i)\right) \cdot w_{L_q}^{\delta_2}\left(\bigwedge_{\substack{1 \leq i \leq m \\ i \text{ even}}} \alpha_{\sigma(h_i)}(a_i)\right) \\
&= v\left(\bigwedge_{i=1}^{m} \alpha_{\sigma(h_i)}(a_i)\right)
\end{aligned}$$

since the NP-continuum functions $w_{L_q}^{\delta_1}, w_{L_q}^{\delta_2}$ satisfy Ax[8], but Ex fails for v since for example for $q = 1$, we have

$$v(\alpha_1(a_1) \wedge \alpha_2(a_2)) = \frac{1}{4}$$

$$v(\alpha_1(a_4) \wedge \alpha_2(a_6)) = \frac{1}{4}(1 - {\delta_2}^2)$$

and these are not equal when $\delta_2 \neq 0$. It follows that a probability function satisfying Ax without Ex exists, so PIP does not imply Ex.

PIP does, however, imply the principles of Predicate Exchangeability, Strong Negation and Variable Exchangeability (and consequently so does INV of course) [36, Chapter

[8] [29, 30] or Theorem 18.2 of [36].

39]. Each of these principles can be represented as invariance under a particular $F \in \mathcal{F}$, so they follow as a result of requiring PIP. For example, if $F(\Theta)$ is the result of permuting some (fixed) relation symbol R_i everywhere in Θ with another relation symbol R_j of the same arity, then F gives Px and is one of the functions from $\mathcal{F}$ since it is easily seen to satisfy conditions (A) and (B).

In addition, we know Super Regularity to be inconsistent with INV for unary languages [34]. Later on in this section, we will resolve whether Super Regularity is consistent with PIP and whether it is consistent with INV for polyadic languages. In Chapter 3 we explore the connections between PIP, Spectrum Exchangeability and Language Invariance, where these principles are also described, and as already alluded to above, the relationship between PIP and Atom Exchangeability will be investigated in detail in Chapter 4.

Firstly however, we continue from where the previous section ended, proceeding as follows. Let w be an arbitrary probability function on SL and define $w' : SL \to [0,1]$ by first setting for state descriptions $\Theta(a_1, \dots, a_n)$, $n \in \mathbb{N}$,

$$w'(\Theta(a_1, \dots, a_n)) = \frac{1}{|\mathcal{F}|} \sum_{F \in \mathcal{F}} w_F(\Theta(a_1, \dots, a_n)) \tag{2.2}$$

where $w_F(\Theta(a_1, \dots, a_n)) = w(F(\Theta)(a_1, \dots, a_n))$.

Lemma 2.3. *The function w' defined in* (2.2) *extends uniquely to a probability function on SL. Moreover,* (2.2) *holds even when the constants $a_1, \dots, a_n$ are replaced by any other distinct constants $b_1, \dots, b_n$.*

Proof. Let $F \in \mathcal{F}$. w_F extends to a probability function on SL since every such F is $\bar{\eta}$ for an automorphism η of BL that permutes state formulae by Theorem 1 of [35], and so the extension of w_F to every $\theta \in SL$ is $w_\eta(\theta) = w(\eta\theta)$ for this η, and w_η is a probability function by the remark on page 19.

Alternatively, we can check that conditions (P1′)-(P3′) from page 13 hold for w_F. w_F clearly satisfies (P1′) and (P2′). To check (P3′) note that by condition (B) on F, for state descriptions $\Phi(a_1, \dots, a_n, a_{n+1})$, $\Theta(a_1, \dots, a_n)$ we have

$$\Phi(a_1, \dots, a_{n+1}) \models \Theta(a_1, \dots, a_n) \iff F(\Phi)(a_1, \dots, a_{n+1}) \models F(\Theta)(a_1, \dots, a_n)$$

since for example, in the first direction, $\Phi[z_1, \ldots, z_n] = \Theta(z_1, \ldots, z_n)$, so when $F \in \mathcal{F}$, $F(\Phi)(a_1, \ldots, a_{n+1})$ logically implies

$$F(\Phi)[z_1, \ldots, z_n](a_1, \ldots, a_n) = F(\Phi[z_1, \ldots, z_n])(a_1, \ldots, a_n) = F(\Theta)(a_1, \ldots, a_n)$$

and similarly in the other direction.

Consequently, since (P3′) holds for w and since $F(\Phi)(a_1, \ldots, a_{n+1})$ run through the state descriptions for $a_1, \ldots, a_{n+1}$ when $\Phi(a_1, \ldots, a_{n+1})$ do so, we have

$$\begin{aligned} w_F(\Theta(a_1, \ldots, a_n)) &= w(F(\Theta)(a_1, \ldots, a_n)) \\ &= \sum_{F(\Phi)(a_1, \ldots, a_{n+1}) \models F(\Theta)(a_1, \ldots, a_n)} w(F(\Phi)(a_1, \ldots, a_{n+1})) \\ &= \sum_{\Phi(a_1, \ldots, a_{n+1}) \models \Theta(a_1, \ldots, a_n)} w_F(\Phi(a_1, \ldots, a_{n+1})). \end{aligned}$$

So (P3′) holds for w_F and hence w_F extends uniquely to a probability function on SL.

w' is therefore a convex combination of probability functions on SL and thus defines a probability function on SL.

The rest of the lemma follows upon noting that any probability function u on SL satisfies

$$u(\Theta(b_1, b_2, \ldots, b_n)) = \sum_{\Phi(a_1, a_2, \ldots, a_k) \models \Theta(b_1, b_2, \ldots, b_n)} u(\Phi(a_1, a_2, \ldots, a_k)),$$

where k is large enough for the $b_1, \ldots, b_n$ to be included amongst $a_1, \ldots, a_k$. □

Proposition 2.4. *The probability function w' defined in (2.2) satisfies PIP. If, in addition, the original probability function w satisfies Ex + SReg then so does w'.*

Proof. To see that w' satisfies PIP, let $\Theta(z_1, \ldots, z_n)$, $\Phi(z_1, \ldots, z_n)$ be state formulae of L with $G(\Theta) = \Phi$ for some $G \in \mathcal{F}$. Consider the set

$$\mathcal{F}' = \{F G^{-1} \mid F \in \mathcal{F}\}.$$

$\mathcal{F}$ is closed under composition and inverse of functions [35], so for every $F \in \mathcal{F}$, we have $F G^{-1} \in \mathcal{F}$ thus $\mathcal{F}' \subseteq \mathcal{F}$. Conversely, every $F \in \mathcal{F}$ can be written as the composition of $\mathsf{F}\,\mathsf{G} \in \mathcal{F}$ and G^{-1}, so $\mathcal{F} \subseteq \mathcal{F}'$. Therefore, the sets $\mathcal{F}$ and $\mathcal{F}'$ are equal and

$$\begin{aligned}
w'(\Theta(b_1,\dots,b_n)) &= \frac{1}{|\mathcal{F}|}\sum_{\mathsf{F}\in\mathcal{F}} w(\mathsf{F}(\Theta)(b_1,\dots,b_n)) \\
&= \frac{1}{|\mathcal{F}|}\sum_{\mathsf{F}\in\mathcal{F}} w(\mathsf{F}\,\mathsf{G}^{-1}\left(\mathsf{G}(\Theta)\right)(b_1,\dots,b_n)) \\
&= \frac{1}{|\mathcal{F}'|}\sum_{\mathsf{F}\,\mathsf{G}^{-1}\in\mathcal{F}'} w(\mathsf{F}\,\mathsf{G}^{-1}(\Phi)(b_1,\dots,b_n)) \\
&= \frac{1}{|\mathcal{F}|}\sum_{\mathsf{F}\in\mathcal{F}} w(\mathsf{F}(\Phi)(b_1,\dots,b_n)) \\
&= w'(\Phi(b_1,\dots,b_n)).
\end{aligned}$$

Suppose w satisfies Ex. Then for every $\mathsf{F} \in \mathcal{F}$ and $b_1,\dots,b_n$, $b'_1,\dots,b'_n$ from the a_i

$$w(\mathsf{F}(\Theta)(b_1,\dots,b_n)) = w(\mathsf{F}(\Theta)(b'_1,\dots,b'_n))$$

so w_{F} satisfies Ex on state descriptions and hence[9] on SL. Consequently, so does w'.

Now suppose w is super regular. The extension of each w_{F} to a probability function on SL is unique and w' is defined as the weighted sum of these extensions. Notice that the permutation that maps each state formula to itself trivially satisfies (A) and (B), so Id $\in \mathcal{F}$.[10] It follows that w must be the extension to SL of w_{Id} defined on state descriptions of L and w is therefore one of the summands in the calculation of w'. So $w'(\theta) \geq \frac{1}{|\mathcal{F}|}\,w(\theta) > 0$ for every consistent $\theta \in SL$ and so w' is super regular. $\square$

The existence of a probability function w' that satisfies PIP, SReg and Ex follows, since a w satisfying Ex and SReg exists, see for example Chapter 26 of [36]. Note also that since PIP implies the principles of Predicate Exchangeability, Strong Negation and Variable Exchangeability, w' will also satisfy these principles.

The consistency of Super Regularity with PIP is interesting due to the restrictive nature of SReg. Yet this consistency becomes perhaps even more noteworthy in view of the fact that INV contradicts SReg, as we shall now show. The case for languages containing only unary predicate symbols follows from the results in [34] that we have

[9] By our remark from page 14.
[10] Id was defined on page 16.

already mentioned and we will extend it to polyadic languages. For simplicity, we shall construct the argument for a binary language, however the result generalises similarly to languages of higher arities.

Let L_1 denote the language with a single unary predicate symbol P and let L be the language with a single binary relation symbol R. Let $\phi \in SL$ be the sentence

$$\forall x\,(\forall y\, R(x,y) \vee \forall y\, \neg R(x,y)).$$

For $\mathcal{M} \in \mathcal{T}L$ such that $\mathcal{M} \models \phi$, define $\beta(\mathcal{M}) \in \mathcal{T}L_1$ by

$$\mathcal{M} \models R(a_i, a_1) \iff \beta(\mathcal{M}) \models P(a_i),$$

so β is a bijection between $\{\mathcal{M} \in \mathcal{T}L \,|\, \mathcal{M} \models \phi\}$ and $\mathcal{T}L_1$.

For $\psi \in SL$, define ψ^* to be the result of replacing each occurrence of $R(t_1, t_2)$ in ψ, where t_1, t_2 are any terms of L, by $P(t_1)$. Then for $\mathcal{M} \models \phi$ it follows easily by induction on the complexity of L-formulae that

$$\mathcal{M} \models \psi \iff \beta(\mathcal{M}) \models \psi^*.$$

Similarly, for $\xi \in SL_1$ we define ξ^+ to be the result of replacing each occurrence of $P(t_1)$ in ξ by $R(t_1, a_1)$. Then for $\mathcal{M} \models \phi$

$$\mathcal{M} \models \xi^+ \iff \beta(\mathcal{M}) \models \xi.$$

In [34] an automorphism[11] δ of BL_1 is specified, with the property

$$\delta[P(a_1) \wedge P(a_2)] = [P(a_1) \wedge P(a_2) \wedge P(a_3)].$$

Using this automorphism δ, define a bijection $\tau : \mathcal{T}L \to \mathcal{T}L$ in the following way:

$$\tau(\mathcal{M}) = \begin{cases} \beta^{-1}(\delta(\beta(\mathcal{M}))) & \text{if } \mathcal{M} \models \phi, \\ \mathcal{M} & \text{otherwise.} \end{cases}$$

Then τ is an automorphism of BL since for $\psi \in SL$,

$$\tau[\psi] = \tau[(\psi \wedge \neg\phi) \vee (\psi \wedge \phi)] = [\psi \wedge \neg\phi] \cup [(\delta(\psi^*))^+ \wedge \phi] = [(\psi \wedge \neg\phi) \vee ((\delta(\psi^*))^+ \wedge \phi)] = [\theta]$$

[11] referred to as γ in [34].

for the sentence $\theta \equiv (\psi \wedge \neg\phi) \vee ((\delta(\psi^*))^+ \wedge \phi)$ and similarly, for every $\theta \in SL$, $\tau^{-1}[\theta] = [\psi]$ for some $\psi \in SL$.

Let $\psi \in SL$ be the sentence $R(a_1, a_1) \wedge R(a_2, a_1)$. Then ψ^* is $P(a_1) \wedge P(a_2)$, $\delta(\psi^*) = P(a_1) \wedge P(a_2) \wedge P(a_3)$, so $(\delta(\psi^*))^+ = R(a_1, a_1) \wedge R(a_2, a_1) \wedge R(a_3, a_1)$. Consequently, for any probability function w satisfying INV and ϕ as above, we require

$$w(R(a_1, a_1) \wedge R(a_2, a_1) \wedge \phi) = w(R(a_1, a_1) \wedge R(a_2, a_1) \wedge R(a_3, a_1) \wedge \phi).$$

On the other hand, $w(R(a_1, a_1) \wedge R(a_2, a_1) \wedge \phi) =$

$$w(R(a_1, a_1) \wedge R(a_2, a_1) \wedge R(a_3, a_1) \wedge \phi) + w(R(a_1, a_1) \wedge R(a_2, a_1) \wedge \neg R(a_3, a_1) \wedge \phi).$$

However then

$$w(R(a_1, a_1) \wedge R(a_2, a_1) \wedge \neg R(a_3, a_1) \wedge \phi) = 0$$

and this sentence is satisfiable. Therefore w cannot satisfy super regularity. We conclude that SReg is inconsistent with INV.

2.4 PIP and Similarity

Previous investigations into the Permutation Invariance Principle have utilised the equivalent Nathanial's Invariance Principle (NIP) [33, 35], which involves the idea of similarity. These two principles have since been unified by [36] under the name PIP, however, the concept of similarity is essential to working with PIP and shall be of use to us in particular in Chapter 4. We present it here.

Definition 2.5. State formulae $\Theta(z_1, \ldots, z_n)$, $\Phi(z_1, \ldots, z_n)$ are said to be *similar*, denoted by[12] $\Theta(\vec{z}) \approx \Phi(\vec{z})$, if for all (distinct) $i_1, \ldots, i_t$ and $j_1, \ldots, j_s$ from $\{1, \ldots, n\}$ and $\sigma : \{z_{i_1}, \ldots, z_{i_t}\} \twoheadrightarrow \{z_{j_1}, \ldots, z_{j_s}\}$, we have

$$\Theta[z_{i_1}, \ldots, z_{i_t}] = (\Theta[z_{j_1}, \ldots, z_{j_s}])_\sigma \iff \Phi[z_{i_1}, \ldots, z_{i_t}] = (\Phi[z_{j_1}, \ldots, z_{j_s}])_\sigma. \tag{2.3}$$

We define two state descriptions to be similar analogously, with $z_1, \ldots, z_n$ replaced by (distinct) constants $b_1, \ldots, b_n$. Note also that in the definition of similarity, if $t = s$ and $\sigma(z_{i_k}) = z_{j_k}$ for each $k \in \{1, \ldots, t\}$, then

[12] We use $\vec{z}$ as a shorthand for the ordered tuple $\langle z_1, \ldots, z_n \rangle$. We will use vector notation as shorthand for other ordered tuples (that will be clear from the context) throughout this account.

$$\Theta[z_{i_1},\dots,z_{i_t}] = \Theta[z_{j_1},\dots,z_{j_t}](z_{i_1}/z_{j_1},\dots,z_{i_t}/z_{j_t}) \iff \Phi[z_{i_1},\dots,z_{i_t}] = \Phi[z_{j_1},\dots,z_{j_t}](z_{i_1}/z_{j_1},\dots,z_{i_t}/z_{j_t})$$

where $\Theta[z_{j_1},\dots,z_{j_t}](z_{i_1}/z_{j_1},\dots,z_{i_t}/z_{j_t})$ is the result of simultaneously replacing each occurrence of z_{j_k} in $\Theta[z_{j_1},\dots,z_{j_t}]$ by z_{i_k}, $k \in \{1,\dots,t\}$. This observation shall be used in what follows without further mention.

Nathanial's Invariance Principle, NIP

For similar state descriptions $\Theta(b_1,\dots,b_n)$ *and* $\Phi(b_1,\dots,b_n)$*,*

$$w(\Theta(b_1,\dots,b_n)) = w(\Phi(b_1,\dots,b_n)).$$

The following theorem [35, Theorem 3] brings together the notions of similarity and automorphisms permuting state formulae (and hence NIP and PIP).

Theorem 2.6. *State formulae* Θ *and* Φ *are similar if and only if there is a permutation of state formulae that satisfies (A) and (B) and maps* Θ *to* Φ*.*

Theorem 2.6 combined with Lemma 2.2 means that the definition of similarity, Definition 2.5, can be simplified considerably. For L an r-ary language, it suffices to consider $t, s \leq r$ in (2.3), as we now show.

Proposition 2.7. *Let* L *be an* r*-ary language. Then Definition 2.5 can be equivalently stated as:*

State formulae $\Theta(z_1,\dots,z_n)$, $\Phi(z_1,\dots,z_n)$ *are* similar *if for all (distinct)* $i_1,\dots,i_t$ *and* $j_1,\dots,j_s$ *from* $\{1,\dots,n\}$ *such that* $t, s \leq r$ *and* $\sigma : \{z_{i_1},\dots,z_{i_t}\} \twoheadrightarrow \{z_{j_1},\dots,z_{j_s}\}$,

$$\Theta[z_{i_1},\dots,z_{i_t}] = (\Theta[z_{j_1},\dots,z_{j_s}])_\sigma \iff \Phi[z_{i_1},\dots,z_{i_t}] = (\Phi[z_{j_1},\dots,z_{j_s}])_\sigma. \tag{2.4}$$

Proof. Definition 2.5 clearly implies (2.4), since choosing *distinct* $i_1,\dots,i_t$ and $j_1,\dots,j_s$ from $\{1,\dots,n\}$ means that $t, s \leq n$, therefore every such choice with $t, s \leq r$ is covered by Definition 2.5 (if $r \leq n$ this is immediate, and if $n < r$ then taking $t, s \leq r$ introduces no new possibilities). We now show that it is also the case that requiring (2.4) to hold implies Definition 2.5.

Suppose not. Then there exist state formulae $\Theta(z_1, \ldots, z_n)$, $\Phi(z_1, \ldots, z_n)$ of L, for which (2.4) holds for all $t, s \leq r$, but (2.4) fails when $s = l$, $t = k$ with $k > r$.[13] In other words,

$$\Theta[z_{i_1}, \ldots, z_{i_k}] = (\Theta[z_{j_1}, \ldots, z_{j_l}])_\sigma$$

say, but

$$\Phi[z_{i_1}, \ldots, z_{i_k}] \neq (\Phi[z_{j_1}, \ldots, z_{j_l}])_\sigma.$$

Recall that $(\Phi[z_{j_1}, \ldots, z_{j_l}])_\sigma$ is the unique state formula $\Psi(z_{i_1}, \ldots, z_{i_k})$ such that $\Psi(\sigma(z_{i_1}), \ldots, \sigma(z_{i_k})) = \Phi[z_{j_1}, \ldots, z_{j_l}]$. So we have $\Phi[z_{i_1}, \ldots, z_{i_k}] \neq \Psi(z_{i_1}, \ldots, z_{i_k})$, and in particular, there must be some relation symbol R_d of L on which $\Phi[z_{i_1}, \ldots, z_{i_k}]$ and $\Psi(z_{i_1}, \ldots, z_{i_k})$ disagree. Suppose

$$\Phi[z_{i_1}, \ldots, z_{i_k}] \models R_d(z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}})$$

and

$$\Psi(z_{i_1}, \ldots, z_{i_k}) \not\models R_d(z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}}),$$

the other case being symmetric.

Let τ denote the restriction of σ to $z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}}$. Let the image of τ consist of the variables $z_{j_{g_1}}, \ldots, z_{j_{g_v}}$, so that $\tau : \{z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}}\} \twoheadrightarrow \{z_{j_{g_1}}, \ldots, z_{j_{g_v}}\}$. Then

$$\Phi[z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}}] \models R_d(z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}})$$

and

$$(\Phi[z_{j_{g_1}}, \ldots, z_{j_{g_v}}])_\tau \not\models R_d(z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}})$$

but $\Theta[z_{i_{h_1}}, \ldots, z_{i_{h_{r_d}}}] = (\Theta[z_{j_{g_1}}, \ldots, z_{j_{g_v}}])_\tau$. However, $r_d \leq r$ and this contradicts our assumption that (2.4) holds for all $t, s \leq r$, so no such k, l exist. Therefore, requiring that (2.4) holds for $t, s \leq r$, implies it holds for all possible t, s, and thus $\Theta(\vec{z}), \Phi(\vec{z})$ are similar according to the original definition. $\square$

The notion of similarity is based on the 'structure' of state formulae. Informally, two state formulae are similar if wherever a sub-state formula repeats in one it also does in the other, and wherever a sub-state formula repeats 'blown up' in one, it does so

[13] We place no restriction on l here (of course other than $l \leq k$), since as shall be shown, it suffices to assume that $k > r$ to arrive at the required contradiction.

in the other. So similar state formulae are those with the same underlying structure, in terms of where their substructures locally repeat or repeat expanded. This gives us a way of checking when two state formulae can be mapped one to the other by a function permuting state formulae and satisfying (A) and (B).

Example. Consider a language containing a single binary relation symbol. The state formulae $\Theta(z_1, z_2, z_3, z_4), \Phi(z_1, z_2, z_3, z_4)$ *represented respectively by the matrices below, are similar:*

$$\begin{matrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{matrix} \qquad\qquad \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}.$$

This can be checked (somewhat laboriously) by confirming that the sub-state formulae of Θ *and* Φ *on* 1 *and* 2 *variables satisfy* (2.3). *For instance,*

$$\Theta[z_1, z_2] \equiv \Theta[z_1, z_3](z_1/z_1, z_2/z_3) \equiv \Theta[z_4, z_3](z_1/z_4, z_2/z_3) \tag{2.5}$$

$$\Phi[z_1, z_2] \equiv \Phi[z_1, z_3](z_1/z_1, z_2/z_3) \equiv \Phi[z_4, z_3](z_1/z_4, z_2/z_3) \tag{2.6}$$

and

$$(\Theta[z_1])_\rho \equiv (\Theta[z_4])_\tau \equiv \Theta[z_1, z_4] \equiv \Theta[z_4, z_1](z_1/z_4, z_4/z_1) \tag{2.7}$$

$$(\Phi[z_1])_\rho \equiv (\Phi[z_4])_\tau \equiv \Phi[z_1, z_4] \equiv \Phi[z_4, z_1](z_1/z_4, z_4/z_1) \tag{2.8}$$

where $\rho(z_1) = \rho(z_4) = z_1$, $\tau(z_1) = \tau(z_4) = z_4$. *Checking* (2.3) *is satisfied by the remaining sub-state formulae on* 1 *and* 2 *variables is done similarly. Notice that by* (2.5), (2.7) *and* (2.6), (2.8) *respectively we also have that* $(\Theta[z_1, z_2])_\sigma \equiv \Theta[z_1, z_3, z_4]$ *and* $(\Phi[z_1, z_2])_\sigma \equiv \Phi[z_1, z_3, z_4]$, *where* σ *sends* z_1 *and* z_4 *to* z_1, *and* z_3 *to* z_2.

Chapter 3

PIP, Spectrum Exchangeability and Language Invariance

3.1 Introduction

Symmetry considerations in Polyadic Inductive Logic have produced two key players to date - the Permutation Invariance Principle - introduced in the previous chapter, and the principle of *Spectrum Exchangeability* (Sx), which we shall explain shortly. In this chapter we explore the relationship between PIP and Sx, thus hoping to elucidate the current polyadic symmetry picture. We will find that these two principles, while originating from entirely different motivations, share some close connections.

In this first section, we present the key principles and surrounding theory that will be used in the chapter. In particular, we describe Sx, the functions $u^{\bar{p},L}$ that satisfy it, the principle of Language Invariance, and the family of probability functions $u^{\bar{p},L}_{\bar{E}}$ that satisfy PIP.

After introducing the required notions, we will show that PIP does not imply Sx; that unary language invariant families with PIP can have multiple extensions to general language invariant families with PIP (in contrast to the situation with Sx), and finally, that we can generate language invariant families with PIP that satisfy Sx up to any given arity, but fail to satisfy it for languages of higher arity. Results from Sections 3.2 and 3.3 appear also in [36, Chapter 42].

Spectrum Exchangeability

Consider the following formulation of Atom Exchangeability (for the unary language L_q), which is equivalent to the formulation given on page 15 *if we assume Ex holds*:

$$w(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)) \textit{ depends only on the multiset } \{m_1, m_2, \ldots, m_{2^q}\},$$
$$\textit{where } m_j \textit{ is the number of times the atom } \alpha_j \textit{ appears amongst the } \alpha_{h_i}.$$

The multiset $\{m_1, m_2, \ldots, m_{2^q}\}$ is known as the *spectrum* of the state description. Since in the unary case knowing which atom a constant satisfies completely determines its behaviour, two constants that satisfy the same atom within a state description are *indistinguishable* from each other with respect to that state description. More formally, b_i and b_j are indistinguishable in $\Theta(b_1, \ldots, b_n)$, denoted by $b_i \sim_\Theta b_j$, if for any predicate symbol P_d of the language

$$\Theta(b_1, \ldots, b_n) \models P_d(b_i) \iff \Theta(b_1, \ldots, b_n) \models P_d(b_j).$$

$\sim_\Theta$ defined in this way is an equivalence relation on the set $\{b_1, \ldots, b_n\}$.

We can extend this notion to a polyadic language L by defining constants b_i and b_j to be indistinguishable in a state description $\Theta(b_1, \ldots, b_n)$, $b_i \sim_\Theta b_j$, if for any relation symbol R_d of L and $b_{k_1}, \ldots, b_{k_u}, b_{k_{u+2}}, \ldots, b_{k_{r_d}}$ from $\{b_1, \ldots, b_n\}$

$$\begin{aligned}\Theta(b_1, \ldots, b_n) &\models R_d(b_{k_1}, \ldots, b_{k_u}, b_i, b_{k_{u+2}}, \ldots, b_{k_{r_d}}) \\ &\iff \Theta(b_1, \ldots, b_n) \models R_d(b_{k_1}, \ldots, b_{k_u}, b_j, b_{k_{u+2}}, \ldots, b_{k_{r_d}}). \quad (3.1)\end{aligned}$$

The *spectrum* of a state description Θ, denoted by $\mathcal{S}(\Theta)$, is the multiset of the sizes of the equivalence classes of $\sim_\Theta$ (written in descending order), and Spectrum Exchangeability[1] is stated as follows:

Spectrum Exchangeability, Sx

For state descriptions $\Theta(b_1, \ldots, b_n)$, $\Phi(b'_1, \ldots, b'_n)$, *if* $\mathcal{S}(\Theta) = \mathcal{S}(\Phi)$, *then*[2]

$$w(\Theta(b_1, \ldots, b_n)) = w(\Phi(b'_1, \ldots, b'_n)).$$

Clearly, Sx is a polyadic generalisation of Ax as stated above. The interested reader may look to [20, 22, 25, 31, 36] for investigations on Sx.

[1] This principle first appeared in [31].

[2] where $b'_1, \ldots, b'_n$ as well as the usual $b_1, \ldots, b_n$ are some distinct choices from the a_i.

At this point it will be useful to introduce the family of functions $u^{\bar{p},L}$. These functions have been prominent in polyadic symmetry, and are investigated for example in [20, 23, 24, 36]. They form the building blocks of probability functions satisfying Sx and they will be used throughout this chapter.

The Functions $u^{\bar{p},L}$

A sequence of colours $\langle c_1, \dots, c_n \rangle \in \{0, 1, 2, \dots\}^n$ is picked at random, where each c_j is picked to be i with probability p_i. At each stage j, we pick a state description $\Theta_j(b_1, \dots, b_j)$ that extends our current state description. However, if $c_j = c_k \neq 0$ for some $k < j$, $\Theta_j(b_1, \dots, b_j)$ must be chosen such that $b_k \sim_{\Theta_j} b_j$. If a new colour or colour 0 is chosen, the new state description Θ_j is chosen at random from those extending the previous one, on the condition that if $c_k = c_l \neq 0$ for some $k, l < j$, then b_k and b_l remain indistinguishable in Θ_j. $u^{\bar{p},L}(\Theta(b_1, \dots, b_n))$ is defined as the sum of the probabilities of choosing $\vec{c}$ and a state description in the manner described above which equals $\Theta(b_1, \dots, b_n)$.

More formally, let $\mathbb{B}$ be the set of sequences of real numbers $\bar{p} = \langle p_0, p_1, p_2, p_3, \dots \rangle$, with $p_i \geq 0$ for all i, $p_1 \geq p_2 \geq p_3 \geq \dots$ and $\sum_{i=1}^{\infty} p_i = 1$. Let $\vec{c} = \langle c_1, c_2, \dots, c_n \rangle \in \{0, 1, 2, \dots, \}^n$. A state description $\Theta(b_1, \dots, b_n)$ is *consistent with* $\vec{c}$ if whenever $c_j = c_k \neq 0$, $b_j \sim_\Theta b_k$. Let $\mathcal{C}(\vec{c}, \vec{b})$ be the set of all state descriptions for $\vec{b} = \langle b_1, b_2, \dots, b_n \rangle$ consistent with $\vec{c}$. Then

$$u^{\bar{p},L}(\Theta(\vec{b})) = \sum_{\substack{\vec{c} \in \{0,1,2,\dots\}^n \\ \Theta \in \mathcal{C}(\vec{c},\vec{b})}} |\mathcal{C}(\vec{c}, \vec{b})|^{-1} \prod_{i=1}^{n} p_{c_i}. \tag{3.2}$$

Language Invariance

Suppose we have found our rational probability function in $w : SL \to [0, 1]$, a probability function defined on the sentences of a language L. It would be unreasonable to assume that we would know from the start, or at all, that L is the only possible language. So we would like to be able to extend the domain of w to any larger language, and to be able to restrict w to act on sentences of smaller languages (thus in effect defining w on all (finite) languages), while maintaining the probabilities w gives to sentences of L.

This concept, now commonly known as *Language Invariance*, has been around since the early days of the subject, forming Carnap's Axiom A11 of his *Axioms of Invariance* [8]. In our context, this principle is stated for unary languages as follows:

Unary Language Invariance, ULi

A probability function w on a unary language L satisfies Unary Language Invariance if there is a family of probability functions $w^{\mathcal{L}}$, one on each (finite) unary language $\mathcal{L}$, satisfying Ex and Px, such that $w^L = w$ and whenever $\mathcal{L} \subseteq \mathcal{L}'$, $w^{\mathcal{L}} = w^{\mathcal{L}'} \restriction S\mathcal{L}$.

Clearly, if $\mathcal{L} \subseteq L$, we have $w^{\mathcal{L}} = w \restriction S\mathcal{L}$ for such a family. Similarly, we define **Language Invariance, Li** , as the corresponding principle for any (not necessarily unary) language L. We say that w *satisfies (U)Li with $\mathcal{P}$* (or (U)Li + $\mathcal{P}$) for some property $\mathcal{P}$ if every member $w^{\mathcal{L}}$ of such a family containing w also satisfies $\mathcal{P}$.

Language Invariance plays a significant role in the relationship between PIP and Sx, as we shall see in Sections 3.3 and 3.4. For this reason, it will be helpful to first understand some of the connections between Language Invariance and Sx. The following results are from [21, 24] and [36, Chapter 32].

Theorem 3.1. [3] *A probability function w satisfies Li + Sx if and only if there is a measure μ on the Borel subsets of $\mathbb{B}$ such that for $\theta \in SL$,*

$$w(\theta) = \int_{\mathbb{B}} u^{\bar{p},L}(\theta)\ d\mu(\bar{p}). \tag{3.3}$$

In addition, if L contains at least one non-unary relation symbol then the language invariant family containing w is unique.

Theorem 3.2. *Let w be a probability function on a unary language L satisfying ULi + Ax. Then there is a measure μ on the Borel subsets of $\mathbb{B}$ such that for $\theta \in SL$*

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

and thus w satisfies Li + Sx.

[3]This representation theorem is similar in style to de Finetti's Representation Theorem, a key result of the subject, which we shall discuss in Chapter 5, page 66.

Lemma 3.3. *Let $\{w^{\mathcal{L}}\}$, $\{v^{\mathcal{L}}\}$ be language invariant families with Sx such that $w^L = v^L$ for every unary language L. Then $w^{\mathcal{L}} = v^{\mathcal{L}}$ for every language $\mathcal{L}$.*

In other words, if two language invariant families with Sx agree on all unary languages, then they agree on *all* languages. We therefore have that a given unary language invariant family with Ax extends uniquely to a language invariant family with Sx, since such an extension exists by Theorem 3.2 and it is unique by Lemma 3.3. In the opposite direction, the restriction of a Li + Sx family to a ULi family with Ax is unique, by the definition of Language Invariance and since Sx on unary languages is equivalent to Ax[4]. It follows that there is a *one-to-one correspondence* between unary language invariant families with Sx and language invariant families with Sx.

The Functions $u^{\bar{p},L}_{\bar{E}}$

We now describe the probability functions $u^{\bar{p},L}_{\bar{E}}$, introduced in[5] [33] and described also in [36, Chapter 42]. They were thought of as building blocks of functions satisfying PIP and the hope is that future research will lead to a representation theorem for all functions satisfying PIP using (some version of) the $u^{\bar{p},L}_{\bar{E}}$. The $u^{\bar{p},L}_{\bar{E}}$ are closely related to the $u^{\bar{p},L}$; they can be viewed as a variant of the $u^{\bar{p},L}$ *with additional structure.*

Let $\mathbb{B}$ be as defined on page 35. Let $\bar{p}$ be some sequence in[6] $\mathbb{B}$, with $p_0 = 0$.

Define $\mathbb{E}_k$ to be the set of equivalence relations $\equiv_k$ on $\{1, 2, 3, \dots\}^k$ for each $k \geq 1$, and $\mathbb{E} \subseteq \mathbb{E}_1 \times \mathbb{E}_2 \times \mathbb{E}_3 \times \dots$ to consist of the sequences of equivalence relations $\bar{E} = \langle \equiv^{\bar{E}}_1, \equiv^{\bar{E}}_2, \equiv^{\bar{E}}_3, \dots \rangle$ that satisfy the following condition:

$$\textit{If } \langle c_1, \dots, c_k \rangle \equiv^{\bar{E}}_k \langle d_1, \dots, d_k \rangle, \textit{ then for any } s_1, \dots, s_m \in \{1, \dots, k\}$$
$$\textit{(not necessarily distinct), } \langle c_{s_1}, \dots, c_{s_m} \rangle \equiv^{\bar{E}}_m \langle d_{s_1}, \dots, d_{s_m} \rangle. \tag{3.4}$$

Let $\bar{E} = \langle \equiv^{\bar{E}}_1, \equiv^{\bar{E}}_2, \equiv^{\bar{E}}_3, \dots \rangle$ be some sequence in $\mathbb{E}$. A sequence of colours $\vec{c} = \langle c_1, \dots, c_n \rangle \in \{1, 2, 3, \dots\}^n$ is picked at random so that each colour c_j in $\vec{c}$ is chosen independently to be i with probability p_i.

[4]Note that this assumes Ex, but since (U)Li implies Ex, Ex indeed holds.

[5]where they were called $u^{\bar{p},\bar{E}}_L$.

[6]This definition is slightly different to the one given in [33], where $\bar{p}$ is defined as a sequence in $\mathbb{B}_0 = \{\langle p_1, p_2, \dots \rangle \mid 0 \leq p_i \leq 1\, \forall i, \sum_{i=1}^{\infty} p_i = 1, p_1 \geq p_2 \geq \dots\}$. Taking $\mathbb{B}$ with every $\bar{p}$ having $p_0 = 0$ gives an equivalent definition of the $u^{\bar{p},L}_{\bar{E}}$ that will be more convenient for our purposes.

We define a binary relation $\sim_k^{\vec{c},\bar{E}}$ on $\{b_1,\dots,b_n\}^k$ for each k, using the equivalence

$$\langle b_{i_1},\dots,b_{i_k}\rangle \sim_k^{\vec{c},\bar{E}} \langle b_{j_1},\dots,b_{j_k}\rangle \iff \langle c_{i_1},\dots,c_{i_k}\rangle \equiv_k^{\bar{E}} \langle c_{j_1},\dots,c_{j_k}\rangle \tag{3.5}$$

according to our chosen $\bar{E}$.

Finally, for each relation symbol R_d of L and each equivalence class A of the equivalence relation $\sim_{r_d}^{\vec{c},\bar{E}}$ (where r_d is the arity of R_d), we pick either

$$\bigwedge_{\langle b_{i_1},\dots,b_{i_{r_d}}\rangle\in A} R_d(b_{i_1},\dots,b_{i_{r_d}}) \quad \text{or} \quad \bigwedge_{\langle b_{i_1},\dots,b_{i_{r_d}}\rangle\in A} \neg R_d(b_{i_1},\dots,b_{i_{r_d}}),$$

each with probability $\frac{1}{2}$. $u_{\bar{E}}^{\bar{p},L}(\Theta(b_1,\dots,b_n))$ is defined as the sum of the probabilities of choosing $\vec{c}$ and a state description in the manner described above, which equals $\Theta(b_1,\dots,b_n)$.

Definition 3.4. A state description $\Theta(b_1,b_2,\dots,b_n)$ is *consistent with $\vec{c}$ under $\bar{E}$*, if for any relation symbol R_d of L (of arity r_d) and any $i_1,\dots,i_{r_d}$ and $j_1,\dots,j_{r_d}$ from $\{1,\dots,n\}$ (not necessarily distinct) such that $\langle c_{i_1},\dots,c_{i_{r_d}}\rangle \equiv_{r_d}^{\bar{E}} \langle c_{j_1},\dots,c_{j_{r_d}}\rangle$, we have

$$\Theta \models R_d(b_{i_1},\dots,b_{i_{r_d}}) \iff \Theta \models R_d(b_{j_1},\dots,b_{j_{r_d}}). \tag{3.6}$$

In other words, $b_{i_1},\dots,b_{i_{r_d}}$ and $b_{j_1},\dots,b_{j_{r_d}}$ 'behave in the same way' in $\Theta(b_1,b_2,\dots,b_n)$. When $b_{i_1},\dots,b_{i_{r_d}}$ and $b_{j_1},\dots,b_{j_{r_d}}$ are distinct constants, (3.6) can be expressed as

$$\Theta[b_{i_1},\dots,b_{i_{r_d}}] \equiv \Theta[b_{j_1},\dots,b_{j_{r_d}}](b_{i_1}/b_{j_1},\dots,b_{i_{r_d}}/b_{j_{r_d}}).$$

We denote the set of all state descriptions for $\vec{b}=\langle b_1,b_2,\dots,b_n\rangle$ consistent with $\vec{c}$ under $\bar{E}$ by $\mathcal{C}_{\bar{E}}(\vec{c},\vec{b})$.

It follows from the process described above that only those $\vec{c}$ for which $\Theta(\vec{b})\in\mathcal{C}_{\bar{E}}(\vec{c},\vec{b})$ could add a non-zero contribution to $u_{\bar{E}}^{\bar{p},L}(\Theta(\vec{b}))$. Furthermore, we can calculate the contribution of such a $\vec{c}$ to $u_{\bar{E}}^{\bar{p},L}(\Theta(\vec{b}))$ as

$$|\mathcal{C}_{\bar{E}}(\vec{c},\vec{b})|^{-1}\left(\prod_{i=1}^{n} p_{c_i}\right) = \frac{1}{2^g}\left(\prod_{i=1}^{n} p_{c_i}\right), \tag{3.7}$$

where g is the sum of the total number of $\sim_{r_d}^{\vec{c},\bar{E}}$-equivalence classes in $\{b_1,\dots,b_n\}^{r_d}$ for each relation symbol R_d of L, $d=1,\dots,q$. In other words, g is the number of equivalence classes for which choices must be made, thus giving 2^g possible state descriptions in $\mathcal{C}_{\bar{E}}(\vec{c},\vec{b})$.

We therefore write $u^{\bar{p},L}_{\bar{E}}(\Theta(\vec{b}))$ as

$$\sum_{\substack{\vec{c}\in\{1,2,\dots\}^n \\ \Theta\in\mathcal{C}_{\bar{E}}(\vec{c},\vec{b})}} |\mathcal{C}_{\bar{E}}(\vec{c},\vec{b})|^{-1}\prod_{i=1}^{n} p_{c_i}. \tag{3.8}$$

For $\bar{p}\in\mathbb{B}$ with $p_0=0$ and $\bar{E}\in\mathbb{E}$, the function $u^{\bar{p},L}_{\bar{E}}$ in (3.8) determines a probability function on SL that satisfies PIP (and Ex). Moreover, the $u^{\bar{p},L}_{\bar{E}}$ form a language invariant family with PIP.[7]

We remark that in the definition of $u^{\bar{p},L}_{\bar{E}}$, we can equivalently define $\vec{c}=\langle c_1,\dots,c_n\rangle$ to be a sequence of colours from $\{0,1,2,\dots\}^n$, in the same way $\vec{c}$ is defined for the $u^{\bar{p},L}$. However, since we require that $p_0=0$ throughout this account, there is no advantage in doing so. For any $\vec{c}$ that contains the colour 0, $\prod_{i=1}^n p_{c_i}=0$, and so for a state description Θ, such a $\vec{c}$ can only add a zero summand to $u^{\bar{p},L}_{\bar{E}}(\Theta)$. We make a convention of omitting zero summands from our probability functions for the rest of this chapter without further mention.

3.2 Probability Functions satisfying PIP without Sx

Having covered the required background, we begin with our task of clarifying the relationship between PIP and Sx. We already know that Sx implies PIP since similar state descriptions share the same spectrum [35, Corollary 4], and thus any probability function satisfying Sx would also give the same probability to any two similar state descriptions. On the other hand, by constructing a counterexample, we now show that the converse of this statement does not hold; there are probability functions that satisfy PIP but not Sx.

[7] [33, Theorems 1, 2], [36, Chapter 42].

Proposition 3.5. *PIP does not imply Sx.*

Proof. Let L be a language containing a single binary relation symbol R. Let $\bar{p} \in \mathbb{B}$ be the sequence $\left\langle 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, \dots \right\rangle$ and let $\bar{E}$ be the sequence of equivalences $\left\langle \equiv_1^{\bar{E}}, \equiv_2^{\bar{E}}, \equiv_3^{\bar{E}}, \dots \right\rangle$ defined by having the following equivalence classes on pairs of colours:

$$\{\langle 1,1\rangle, \langle 3,3\rangle\} \quad \{\langle 2,2\rangle, \langle 4,4\rangle\} \quad \{\langle 1,2\rangle, \langle 3,4\rangle\} \quad \{\langle 2,1\rangle, \langle 4,3\rangle\}$$

$$\{\langle 1,3\rangle\} \quad \{\langle 3,1\rangle\} \quad \{\langle 1,4\rangle\} \quad \{\langle 4,1\rangle\} \quad \{\langle 2,3\rangle\} \quad \{\langle 3,2\rangle\} \quad \{\langle 2,4\rangle\} \quad \{\langle 4,2\rangle\} \tag{3.9}$$

and satisfying (3.4). So for this $\bar{E}$ we have $\langle 1,2\rangle \equiv_2^{\bar{E}} \langle 3,4\rangle$, $\langle 2,1\rangle \equiv_2^{\bar{E}} \langle 4,3\rangle$, $\langle 1,1\rangle \equiv_2^{\bar{E}} \langle 3,3\rangle$ and $\langle 2,2\rangle \equiv_2^{\bar{E}} \langle 4,4\rangle$, and for all other pairs[8] $\langle c,d\rangle \equiv_2^{\bar{E}} \langle c,d\rangle$ only.

Let $\Theta(b_1,b_2,b_3,b_4)$ be the state description of L represented by the matrix

$$\begin{matrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{matrix} \, . \tag{3.10}$$

Since none of b_1,b_2,b_3,b_4 are pairwise indistinguishable in Θ (if two constants were indistinguishable we would have two identical rows and two identical columns in the matrix representation (3.10)), the spectrum of this state description, $\mathcal{S}(\Theta)$, is $\{1,1,1,1\}$.

Consider $u_{\bar{E}}^{\bar{p},L}(\Theta(b_1,b_2,b_3,b_4))$. We choose a sequence $\langle c_1,c_2,c_3,c_4\rangle$ of colours, so that each c_j is chosen independently to be one of $\{1,2,3,4\}$ with probability $\frac{1}{4}$, and every other colour with probability 0.

Firstly, note that Θ is not consistent under $\bar{E}$ with any $\vec{c}$ in which a colour appears more than once. To see this, let $\vec{c} = \langle c_1,c_2,c_3,c_4\rangle$ and suppose $c_k = c_l$ for some $k,l \in \{1,2,3,4\}$, $k \neq l$. Then we have that $\langle c_k,c_j\rangle \equiv_2^{\bar{E}} \langle c_l,c_j\rangle$ for any $j \in \{1,2,3,4\}$, so for Θ to be consistent with this $\vec{c}$ under $\bar{E}$, we must have $\Theta \models R(b_k,b_j) \iff \Theta \models R(b_l,b_j)$ for every j. This means there would be two identical rows in the matrix representing Θ, which is not the case.

Next we examine the case where each of the four colours is selected exactly once. Let $\vec{c} = \langle 1,2,3,4\rangle$ say. Since $\langle 1,2\rangle \equiv_2^{\bar{E}} \langle 3,4\rangle$, for Θ to be consistent with this $\vec{c}$ under $\bar{E}$

[8] By (3.4) we also have that $1 \equiv_1^{\bar{E}} 3$ and $2 \equiv_1^{\bar{E}} 4$, and the equivalence is preserved upwards too, so for example $\langle 1,2,1\rangle \equiv_3^{\bar{E}} \langle 3,4,3\rangle$ and so on, but we are focusing on pairs since the language is binary.

we require that

$$\Theta[b_1, b_2] \equiv \Theta[b_3, b_4](b_1/b_3, b_2/b_4).$$

However $\Theta[b_1, b_2]$ is

$$\begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix}$$

and $\Theta[b_3, b_4](b_1/b_3, b_2/b_4)$ is

$$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$$

so Θ is not consistent with $\langle 1, 2, 3, 4\rangle$ under $\bar{E}$. Since *none* of the 2×2 submatrices of (3.10) repeat, a similar argument applies to any permutation of the order in which the four colours are picked. Therefore, Θ is consistent with *no* $\vec{c}$ under $\bar{E}$, and hence $u^{\bar{p},L}_{\bar{E}}(\Theta(b_1, b_2, b_3, b_4)) = 0$.

On the other hand, consider the state description $\Phi(b_1, b_2, b_3, b_4)$ of L, represented by the matrix

$$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} . \tag{3.11}$$

None of b_1, b_2, b_3, b_4 are pairwise indistinguishable in Φ, so $\mathcal{S}(\Phi) = \{1, 1, 1, 1\} = \mathcal{S}(\Theta)$.

We now look at $u^{\bar{p},L}_{\bar{E}}(\Phi(b_1, b_2, b_3, b_4))$. Arguing as we did for Θ, we see that no $\vec{c}$ *with a repeated colour* is consistent with Φ under $\bar{E}$ either. However, let $\vec{c} = \langle 1, 2, 3, 4\rangle$. In this case,

$$\Phi[b_1, b_2] = \Phi[b_3, b_4](b_1/b_3, b_2/b_4),$$

$$\Phi[b_2, b_1] = \Phi[b_4, b_3](b_2/b_4, b_1/b_3),$$

and $\Phi[b_1] = \Phi[b_3](b_1/b_3)$, $\Phi[b_2] = \Phi[b_4](b_2/b_4)$,[9] so Φ is consistent with this $\vec{c}$ under $\bar{E}$.

We can calculate the contribution of $\vec{c} = \langle 1, 2, 3, 4\rangle$ to $u^{\bar{p},L}_{\bar{E}}(\Phi)$ by $\left(\prod_{i=1}^4 p_{c_i}\right) \frac{1}{2^g}$, where g is the number of equivalence classes in $\{b_1, b_2, b_3, b_4\}^2$ with respect to the equivalence $\sim_2^{\vec{c},\bar{E}}$, displayed in (3.9). So in this case $g = 12$ and $\left(\prod_{i=1}^4 p_{c_i}\right) \frac{1}{2^g} = \left(\frac{1}{4}\right)^4 \frac{1}{2^{12}} =$

[9]Equivalently, $\Phi \models R(b_1, b_1) \iff \Phi \models R(b_3, b_3)$ and $\Phi \models R(b_2, b_2) \iff \Phi \models R(b_4, b_4)$ both hold.

$\frac{1}{2^{20}}$. Furthermore, due to the heavily symmetric structure of Φ, any choice of $\vec{c} = \langle c_1, c_2, c_3, c_4\rangle$ where c_1, c_2, c_3, c_4 are all distinct would contribute the same non-zero factor to $u^{\bar{p},L}_{\bar{E}}(\Phi)$ by a similar argument to the one above. Since there are 24 possible permutations σ of $\{1,2,3,4\}$, and it can be checked that Φ is consistent under $\bar{E}$ with $\langle\sigma(1),\sigma(2),\sigma(3),\sigma(4)\rangle$ for each of them, the probability $u^{\bar{p},L}_{\bar{E}}(\Phi(b_1,b_2,b_3,b_4)) = \frac{3}{2^{17}}$.

We conclude that

$$u^{\bar{p},L}_{\bar{E}}(\Phi(b_1,b_2,b_3,b_4)) > u^{\bar{p},L}_{\bar{E}}(\Theta(b_1,b_2,b_3,b_4))$$

while $\mathcal{S}(\Phi) = \mathcal{S}(\Theta)$. So for $\bar{p}, \bar{E}, L$ as above, $u^{\bar{p},L}_{\bar{E}}$ satisfies PIP but does not satisfy Sx. $\square$

We remark that it is true in general that for a state description Θ consistent with $\vec{c}$ under $\bar{E}$, if $\vec{c}$ is such that $c_k = c_l$ then $b_k \sim_\Theta b_l$. This is the case since if $j_1, \dots, j_{r_d}$ are formed from $i_1, \dots, i_{r_d}$ by swapping occurrences of k and l, then $\langle c_{i_1}, \dots, c_{i_{r_d}}\rangle \equiv^{\bar{E}}_{r_d} \langle c_{j_1}, \dots, c_{j_{r_d}}\rangle$ because $\langle c_{i_1}, \dots, c_{i_{r_d}}\rangle = \langle c_{j_1}, \dots, c_{j_{r_d}}\rangle$, and Θ satisfies (3.6).

Secondly, by condition (3.4) on $\bar{E}$, since for any k, $\equiv^{\bar{E}}_k$ defines $\equiv^{\bar{E}}_m$-equivalences for all $m < k$, if $\langle c_1, \dots, c_k\rangle \equiv^{\bar{E}}_k \langle d_1, \dots, d_k\rangle$ but $\langle c_1, \dots, c_k\rangle \neq \langle d_1, \dots, d_k\rangle$, then we must have $c_j \equiv^{\bar{E}}_1 d_j$ but $c_j \neq d_j$ for at least one j from $\{1, \dots, k\}$.

3.3 PIP and Language Invariance

We show that ULi families with PIP can have multiple extensions to Li families with PIP, unlike unary and polyadic language invariant families with Sx (cf. page 37). Our method will be as follows. We will first point out that the $u^{\bar{q},\mathcal{L}}$, $\bar{q} \in \mathbb{B}$ provide one extension from a ULi + PIP to a Li + PIP family. We will then show that for some choice of $\bar{q}, \bar{p}, \bar{E}$ the functions $u^{\bar{q},\mathcal{L}}$ and $u^{\bar{p},\mathcal{L}}_{\bar{E}}$ agree on unary languages but differ on binary languages, and hence conclude that the ULi + PIP family $u^{\bar{q},\mathcal{L}}$ extends to two distinct Li + PIP families for polyadic $\mathcal{L}$, one being $u^{\bar{q},\mathcal{L}}$ and the other $u^{\bar{p},\mathcal{L}}_{\bar{E}}$.

Firstly, notice that ULi + Ax, ULi + Sx and ULi + PIP are all equivalent[10], so the $u^{\bar{q},\mathcal{L}}$ (and any convex combination of the $u^{\bar{q},\mathcal{L}}$ as in (3.3)) on unary $\mathcal{L}$ satisfy ULi +

[10]Recall that PIP is equivalent to Ax on unary languages, page 23.

PIP. Moreover, since (Li +) Sx implies (Li +) PIP, the $u^{\bar{q},\mathcal{L}}$ satisfy Li + PIP. So we have one extension from a ULi + PIP family to a Li + PIP family in the form of the $u^{\bar{q},\mathcal{L}}$.

Now consider the probability function $u^{\bar{p},L}_{\bar{E}}$ used in the proof of Proposition 3.5, where $\bar{p}$ was $\langle 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, \dots \rangle$ and $\bar{E}$ contained the equivalence $\langle 1, 2\rangle \equiv_2^{\bar{E}} \langle 3, 4\rangle$ and all equivalences that follow from condition (3.4). We prove that when the language is unary, this probability function is equal to $u^{\bar{q},L}$ with $\bar{q} = \langle 0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, \dots \rangle \in \mathbb{B}$.

We have $q_1 = p_1 + p_3$, $q_2 = p_2 + p_4$, so when the language is unary, colours 1 and 3 act as if they are one colour, as do colours 2 and 4. To see this, let $\vec{c} \in \{1, 2\}^n$ and $\vec{d} \in \{1, 2, 3, 4\}^n$. Let $D_{\vec{c}}$ contain all $\vec{d}$ formed from this particular $\vec{c}$ by replacing every 1 in $\vec{c}$ by 1 or 3, and every 2 in $\vec{c}$ by 2 or 4. Then the state descriptions for $\vec{b} = \langle b_1, \dots, b_n\rangle$ consistent with this $\vec{c}$ are exactly those consistent with a $\vec{d}$ from $D_{\vec{c}}$ under $\bar{E}$, since colours (1 and 3) and colours (2 and 4) are equivalent under $\bar{E}$. That is, $\mathcal{C}(\vec{c}, \vec{b}) = \mathcal{C}_{\bar{E}}(\vec{d}, \vec{b})$ for the $\vec{d}$ from $D_{\vec{c}}$.

Furthermore, let $A(1) = \{1, 3\}$, $A(2) = \{2, 4\}$. By the definition of $\bar{p}, \bar{q}$,

$$q_{c_i} = \sum_{d \in A(c_i)} p_d.$$

So the probability of picking a particular $\vec{c}$ is

$$\prod_{i=1}^n q_{c_i} = \prod_{i=1}^n \sum_{d \in A(c_i)} p_d = \sum_{\vec{d}: d_i \in A(c_i)} \prod_{i=1}^n p_{d_i}, \tag{3.12}$$

and the $\vec{d}$ such that $d_i \in A(c_i)$ are precisely the $\vec{d}$ in $D_{\vec{c}}$ by definition.

Let $\Theta(b_1, \dots, b_n)$ be a state description in a unary language L. Then using (3.2) and (3.8)

$$u^{\bar{q},L}(\Theta(\vec{b})) = \sum_{\substack{\vec{c} \in \{1,2\}^n \\ \Theta \in \mathcal{C}(\vec{c}, \vec{b})}} |\mathcal{C}(\vec{c}, \vec{b})|^{-1} \prod_{i=1}^n q_{c_i} = \sum_{\substack{\vec{c} \in \{1,2\}^n \\ \Theta \in \mathcal{C}(\vec{c}, \vec{b})}} \sum_{\vec{d} \in D_{\vec{c}}} |\mathcal{C}_{\bar{E}}(\vec{d}, \vec{b})|^{-1} \prod_{i=1}^n p_{d_i} = u^{\bar{p},L}_{\bar{E}}(\Theta(\vec{b}))$$

since $\{\vec{d} \in D_{\vec{c}} \,:\, \vec{c} \in \{1, 2\}^n\} = \{1, 2, 3, 4\}^n$ and since $\Theta \in \mathcal{C}(\vec{c}, \vec{b}) \iff \Theta \in \mathcal{C}_{\bar{E}}(\vec{d}, \vec{b})$ for $d \in D_{\vec{c}}$.

However, in the proof of Proposition 3.5 we showed that $u^{\bar{p},L}_{\bar{E}}$ does not satisfy Sx on binary languages, and therefore cannot be equal to $u^{\bar{q},L}$ when L is binary. We conclude that the language invariant family with PIP $u^{\bar{q},\mathcal{L}}$ extends to (at least) two different language invariant families with PIP. One of these is the language invariant family with Sx, $u^{\bar{q},\mathcal{L}}$, and the other the family $u_{\bar{E}}^{\bar{p},\mathcal{L}}$, where $\bar{q},\bar{p},\bar{E}$ are defined as above.

In fact, by modifying the above argument slightly, we can show that ULi + PIP functions of the form $u^{\bar{q},L}$ with $q_0 = 0$ always have multiple extensions to Li + PIP families.

Proposition 3.6. *Let L be a unary language and let $u^{\bar{q},L}$ be such that $\bar{q} \in \mathbb{B}$ and $q_0 = 0$. Then $u^{\bar{q},L}$ has more than one extension to a language invariant family with PIP.*

Proof. Since $q_0 = 0$, $\bar{q}$ will contain at least one non-zero entry q_1. We construct $\bar{p}$ with $p_0 = 0$ containing (at least) four non-zero entries by splitting q_1 into $p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4} \neq 0$, $p_{i_1} + p_{i_2} + p_{i_3} + p_{i_4} = q_1$. If $\bar{q}$ contains other non-zero entries q_2, q_3 etc, they are added to $\bar{p}$ and ordered such that $p_1 \geq p_2 \geq \dots$. Let $\bar{E}$ contain the equivalence $i_1 \equiv_1^{\bar{E}} i_2 \equiv_1^{\bar{E}} i_3 \equiv_1^{\bar{E}} i_4$ and satisfy condition (3.4). Then $u_{\bar{E}}^{\bar{p},L}$ behaves as $u^{\bar{q},L}$ on unary languages, since similarly to above, in that case colours i_1, i_2, i_3, i_4 behave as colour 1.

On the other hand, $u^{\bar{q},L}$ and $u_{\bar{E}}^{\bar{p},L}$ do not agree on binary languages. To see this, let L contain a single binary relation symbol and let $\Phi(b_1, b_2, b_3, b_4)$ be the state description represented by (3.11). Then if Φ is consistent with some colour vector $\vec{c}$, it will also be consistent under $\bar{E}$ with every $\vec{d}$ formed from $\vec{c}$ by replacing 1 with i_1, i_2, i_3 or i_4. However in addition, it will also be consistent under $\bar{E}$ with every $\vec{d}$ that contains four distinct colours, since if $\langle d_{j_1}, d_{j_2}\rangle \equiv_2^{\bar{E}} \langle d_{k_1}, d_{k_2}\rangle$ then $\Phi[b_{j_1}, b_{j_2}] = \Phi[b_{k_1}, b_{k_2}](b_{j_1}/b_{k_1}, b_{j_2}/b_{k_2})$ since all the 2×2 submatrices of Φ are equal. So $u_{\bar{E}}^{\bar{p},L}(\Phi) > u^{\bar{q},L}(\Phi)$ and therefore $u_{\bar{E}}^{\bar{p},L}$ and $u^{\bar{q},L}$ provide different extensions to Li + PIP for the unary family $u^{\bar{q},L}$. $\square$

3.4 The $u_{\bar{E}}^{\bar{p},L}$ Families and Sx

In the previous section we saw that Li families with PIP and Li families with Sx may agree on unary languages and differ on binary languages. In this section, we investigate this relationship further, and find that in fact, there are Li families with PIP that not only satisfy Sx up to any given arity r, but such that Sx fails for any language of arity higher than r. We will first show this for a family where Sx is only satisfied by unary languages and fails for languages that are binary or higher, and then prove the result for any arity $r \geq 2$.

Let $\bar{p} \in \mathbb{B}$ be such that $p_0 = 0$ and $p_i > 0$ for infinitely many $i \in \mathbb{N}^+$. Define $\bar{E}$ as follows:

- Fix two equivalence classes under $\equiv_1^{\bar{E}}$, one containing all odd colours and one of all even colours:
$$1 \equiv_1^{\bar{E}} 3 \equiv_1^{\bar{E}} 5 \equiv_1^{\bar{E}} \dots ,$$
$$2 \equiv_1^{\bar{E}} 4 \equiv_1^{\bar{E}} 6 \equiv_1^{\bar{E}} \dots .$$

- Add the equivalences $\langle c_1, c_2 \rangle \equiv_2^{\bar{E}} \langle c_2, c_1 \rangle \iff c_1 \equiv_1^{\bar{E}} c_2$.

- For $m > 1$, include every equivalence that must hold by condition (3.4), which in this case amounts to:
$$\textit{If } \langle c_1, c_2 \rangle \equiv_2^{\bar{E}} \langle d_1, d_2 \rangle \textit{ then for } s_1, s_2, \dots, s_m \in \{1, 2\},$$
$$\langle c_{s_1}, c_{s_2}, \dots, c_{s_m} \rangle \equiv_m^{\bar{E}} \langle d_{s_1}, d_{s_2}, \dots, d_{s_m} \rangle, \tag{3.13}$$
where either $c_1 = c_2$, $d_1 = d_2$ and $s_1 = s_2 = \dots = s_m$, or $c_1 \equiv_1^{\bar{E}} c_2 \equiv_1^{\bar{E}} d_1 \equiv_1^{\bar{E}} d_2$ (but they are not equal). This is because every equivalence on pairs either follows by condition (3.4) from an $\equiv_1^{\bar{E}}$-equivalences or it is $\langle c_1, c_2 \rangle \equiv_2^{\bar{E}} \langle c_2, c_1 \rangle$, and in turn, every $\equiv_m^{\bar{E}}$-equivalence follows from one of these.

Proposition 3.7. *The language invariant family $u_{\bar{E}}^{\bar{p},\mathcal{L}}$ with $\bar{p}, \bar{E}$ as above satisfies Sx* only *when the language $\mathcal{L}$ is unary.*

Proof. We first show that when L is unary, this function $u_{\bar{E}}^{\bar{p},L}$ is equal to $u^{\bar{q},L}$ for some $\bar{q} \in \mathbb{B}$ and hence satisfies Sx, and then prove that Sx fails for every r-ary language, $r > 1$.

Let $\bar{q} = \langle 0,\, q_1, q_2, 0, 0, \dots \rangle \in \mathbb{B}$, with
$$q_1 = \sum_{i \text{ odd}} p_i \ , \qquad\qquad q_2 = \sum_{i \text{ even}} p_i \ .$$

Let L be a unary language and let $\Theta(b_1, \dots, b_n)$ be a state description of L. Then[11]

$$u^{\bar{q},L}(\Theta(\vec{b})) = \sum_{\substack{\vec{c}\in\{1,2\}^n \\ \Theta\in\mathcal{C}^L(\vec{c},\vec{b})}} |\mathcal{C}^L(\vec{c},\vec{b})|^{-1} \prod_{i=1}^n q_{c_i}\,,$$

$$u^{\bar{p},L}_{\bar{E}}(\Theta(\vec{b})) = \sum_{\substack{\vec{d}\in\{1,2,3,\dots\}^n \\ \Theta\in\mathcal{C}^L_{\bar{E}}(\vec{d},\vec{b})}} |\mathcal{C}^L_{\bar{E}}(\vec{d},\vec{b})|^{-1} \prod_{i=1}^n p_{d_i}\,.$$

For $\vec{c} \in \{1,2\}^n$, let $D_{\vec{c}}$ denote the set of all those $\vec{d} \in \{1,2,3,\dots\}^n$ obtained from $\vec{c}$ by replacing each occurrence of 1 in $\vec{c}$ by any odd number, and each occurrence of 2 in $\vec{c}$ by any even number. Then the $D_{\vec{c}}$ partition $\{1,2,3,\dots\}^n$. Each $\vec{d} \in D_{\vec{c}}$ is such that $d_i \equiv_1^{\bar{E}} d_j \iff c_i = c_j$. So $\mathcal{C}^L(\vec{c},\vec{b}) = \mathcal{C}^L_{\bar{E}}(\vec{d},\vec{b})$ for $\vec{d} \in D_{\vec{c}}$, since the state descriptions for $\vec{b}$ consistent with $\vec{c}$ are precisely those consistent with the $\vec{d}$ from $D_{\vec{c}}$ under $\bar{E}$.

Furthermore, following the same argument as on page 43,

$$\prod_{i=1}^n q_{c_i} = \sum_{\vec{d}\in D_{\vec{c}}} \prod_{i=1}^n p_{d_i},$$

and so

$$u^{\bar{q},L}(\Theta(\vec{b})) = \sum_{\substack{\vec{c}\in\{1,2\}^n \\ \Theta\in\mathcal{C}^L(\vec{c},\vec{b})}} \sum_{\vec{d}\in D_{\vec{c}}} |\mathcal{C}^L_{\bar{E}}(\vec{d},\vec{b})|^{-1} \prod_{i=1}^n p_{d_i} = u^{\bar{p},L}_{\bar{E}}(\Theta(\vec{b})).$$

Therefore, when the language is unary, $u^{\bar{q},L}$ and $u^{\bar{p},L}_{\bar{E}}$ are equal and hence $u^{\bar{p},L}_{\bar{E}}$ satisfies Sx.

We now show that Sx fails for non-unary languages. Let L^+ contain an r-ary relation symbol R, $r > 1$. Let $\Theta(b_1, b_2, b_3)$ be the state description of L^+ such that

$$\Theta \models R(b_{i_1}, \dots, b_{i_r})$$

whenever $i_1 = i_2$,

$$\Theta \models \neg R(b_{i_1}, \dots, b_{i_r})$$

otherwise. If L^+ contains any other relation symbols, we may just assume they only occur positively in Θ.

Then $\mathcal{S}(\Theta) = \{1,1,1\}$, since if $b_i \sim_\Theta b_j$, we would require using (3.1) that

$$\Theta \models R(b_i, b_i, b_{i_3}, \dots, b_{i_r}) \iff \Theta \models R(b_i, b_j, b_{i_3}, \dots, b_{i_r}),$$

[11] We add a superscript L to $\mathcal{C}(\vec{c},\vec{b})$ and $\mathcal{C}_{\bar{E}}(\vec{d},\vec{b})$ in this section to emphasise the language involved.

but $\Theta \models R(b_i, b_i, b_{i_3}, \ldots, b_{i_r}) \wedge \neg R(b_i, b_j, b_{i_3}, \ldots, b_{i_r})$ whenever $i \neq j$, and thus no b_i, b_j can be indistinguishable in Θ.

Consider $u^{\bar{p},L^+}_{\bar{E}}(\Theta(b_1, b_2, b_3))$. While Θ is consistent with no $\vec{d}$ under $\bar{E}$ in which a colour appears more than once, Θ is consistent, for example[12], with $\langle d_1, d_2, d_3\rangle = \langle 1, 2, 3\rangle$. We have $1 \equiv^{\bar{E}}_1 3$ so by (3.13), $\langle d_1, d_1, \ldots, d_1\rangle \equiv^{\bar{E}}_r \langle d_3, d_3, \ldots, d_3\rangle$, and

$$\Theta \models R(b_1, b_1, \ldots, b_1) \iff \Theta \models R(b_3, b_3, \ldots, b_3)$$

holds. In addition, $\langle d_1, d_3\rangle \equiv^{\bar{E}}_2 \langle d_3, d_1\rangle$, so we require that for $d_{i_1}, \ldots, d_{i_r}$ and $d_{j_1}, \ldots, d_{j_r}$ where for each $g = 1, \ldots, r$ either $i_g = 1$ and $j_g = 3$ or $i_g = 3$ and $j_g = 1$,

$$\Theta \models R(b_{i_1}, \ldots, b_{i_r}) \iff \Theta \models R(b_{j_1}, \ldots, b_{j_r}).$$

But this holds too, since we have either $i_1 = i_2$ and $j_1 = j_2$, or $i_1 \neq i_2$ and $j_1 \neq j_2$. Therefore, since in addition $\prod_{i=1}^3 p_i > 0$, $u^{\bar{p},L^+}_{\bar{E}}(\Theta) > 0$.

On the other hand, let $\Psi(b_1, b_2, b_3)$ be the state description of L^+ such that

$$\Psi \models R(b_1, b_{i_2}, \ldots, b_{i_r}) \wedge R(b_{i_1}, b_2, b_{i_3}, \ldots, b_{i_r})$$

for $i_1, \ldots, i_r \in \{1, 2, 3\}$ and

$$\Psi \models \neg R(b_{i_1}, \ldots, b_{i_r})$$

otherwise. We again assume that any other relation symbols only occur positively in Ψ. Then $\mathcal{S}(\Psi) = \{1, 1, 1\}$ as b_1 cannot be indistinguishable from any other constant because

$$\Psi \models R(b_1, b_1, \ldots, b_1) \wedge \neg R(b_s, b_1, \ldots, b_1)$$

for $s = 2, 3$, and b_2 cannot be indistinguishable from b_3 similarly.

In contrast to Θ, when considering $u^{\bar{p},L^+}_{\bar{E}}(\Psi)$ we find that Ψ is consistent with *no* $\vec{d}$ under $\bar{E}$. Clearly, Ψ is consistent with no $\vec{d}$ that contains a repeated colour. So suppose that every colour in $\vec{d} = \langle d_1, d_2, d_3\rangle$ appears exactly once. By the definition of $\bar{E}$, there must be two colours in $\vec{d}$ that are equivalent under $\equiv^{\bar{E}}_1$ since there are only two equivalence classes under $\equiv^{\bar{E}}_1$; say $d_s \equiv^{\bar{E}}_1 d_t$. We show that for every choice of $s \neq t$ from $\{1, 2, 3\}$ we have $\Psi \notin C^{L^+}_{\bar{E}}(\vec{d}, \vec{b})$. If $d_s \equiv^{\bar{E}}_1 d_3$, then $\langle d_s, d_s, \ldots, d_s\rangle \equiv^{\bar{E}}_r \langle d_3, d_3, \ldots, d_3\rangle$,

[12] We remark that Θ is consistent under $\bar{E}$ with *every* $\vec{d} \in \{1, 2, 3, \ldots\}^3$ in which each colour appears exactly once by a similar argument.

but $\Psi \models R(b_s, b_s, \ldots, b_s) \wedge \neg R(b_3, b_3, \ldots, b_3)$ for $s = 1, 2$. So the remaining possibility is if $d_1 \equiv_1^{\bar{E}} d_2$. Then by the definition of $\bar{E}$ we have $\langle d_1, d_2 \rangle \equiv_2^{\bar{E}} \langle d_2, d_1 \rangle$, and so for any $i_1, \ldots, i_r$ and $j_1, \ldots, j_r$ such that either $i_g = 1$ and $j_g = 2$ or $i_g = 2$ and $j_g = 1$, we require

$$\Psi \models R(b_{i_1}, \ldots, b_{i_r}) \iff \Psi \models R(b_{j_1}, \ldots, b_{j_r}).$$

Let $i_1 = 1, i_2 = 2$, so $j_1 = 2, j_2 = 1$. Then we have

$$\langle d_1, d_2, d_{i_3}, \ldots, d_{i_r} \rangle \equiv_r^{\bar{E}} \langle d_2, d_1, d_{j_3}, \ldots, d_{j_r} \rangle$$

but

$$\Psi \models R(b_1, b_2, b_{i_3}, \ldots, b_{i_r}) \wedge \neg R(b_2, b_1, b_{j_3}, \ldots, b_{j_r}).$$

So Ψ cannot be consistent with $\langle d_1, d_2, d_3 \rangle$ whichever two of d_1, d_2, d_3 are $\equiv_1^{\bar{E}}$-equivalent, and hence $u_{\bar{E}}^{\bar{p}, L^+}(\Psi) = 0$.

Therefore, we have shown that for any r-ary language L^+ with $r \geq 2$, we can find state descriptions with the same spectrum that get different probabilities by $u_{\bar{E}}^{\bar{p}, L^+}$, hence $u_{\bar{E}}^{\bar{p}, L^+}$ does not satisfy Sx. $\square$

We can generalise this method to construct Li + PIP families with Sx holding only for languages of arity at most r.

Theorem 3.8. *There exist language invariant families of probability functions with PIP that satisfy Sx up to any given arity $r \geq 2$, and such that Sx fails on languages of arity higher than r.*

Proof. We provide a method to generate such families. Let $\bar{p} \in \mathbb{B}$ be such that $p_0 = 0$ and infinitely many of the other p_i are non-zero. Define $\bar{E}$ as follows: Fix the $\equiv_1^{\bar{E}}$-equivalences, so that each colour $i \neq 0$ is in one of $r + 1$ many $\equiv_1^{\bar{E}}$-equivalence classes. For each m such that $1 < m < r + 1$, include in $\bar{E}$ every $\equiv_m$-equivalence that does not alter $\equiv_1^{\bar{E}}$, that is

$$\langle c_{i_1}, \ldots, c_{i_m} \rangle \equiv_m^{\bar{E}} \langle d_{i_1}, \ldots, d_{i_m} \rangle \iff (c_{i_1} \equiv_1^{\bar{E}} d_{i_1}) \wedge (c_{i_2} \equiv_1^{\bar{E}} d_{i_2}) \wedge \cdots \wedge (c_{i_m} \equiv_1^{\bar{E}} d_{i_m}).$$

Now define $\equiv_m^{\bar{E}}$ for $m \geq r + 1$ by (the usual condition on $\bar{E}$):

$$\textit{If } \langle c_1, \ldots, c_k \rangle \equiv_k^{\bar{E}} \langle d_1, \ldots, d_k \rangle, \textit{ then for } s_1, \ldots, s_m \in \{1, \ldots, k\}$$
$$\text{(not necessarily distinct)}, \langle c_{s_1}, \ldots, c_{s_m} \rangle \equiv_m^{\bar{E}} \langle d_{s_1}, \ldots, d_{s_m} \rangle. \tag{3.14}$$

Define $\bar{q} \in \mathbb{B}$ with $q_0 = 0$ by:

$$q_f \text{ is the sum of all the } p_i \text{ equivalent to each other under } \equiv_1^{\bar{E}},$$

$f = 1, 2, \ldots$. So $\bar{q}$ has exactly $r+1$ non-zero entries. Then $u_{\bar{E}}^{\bar{p},\mathcal{L}}$ and $u^{\bar{q},\mathcal{L}}$ agree on languages up to highest arity r, following an argument similar to the one presented for the unary case above. It follows that this Li + PIP family $u_{\bar{E}}^{\bar{p},\mathcal{L}}$ satisfies Sx on languages of arity at most r.

To see that $u_{\bar{E}}^{\bar{p},\mathcal{L}}$ does not satisfy Sx on languages of arity higher than r, let L be a language containing an s-ary relation symbol R, $s \geq r+1$. Let $\Theta(b_1, \ldots, b_r, b_{r+1}, b_{r+2})$ be the state description of L defined by

$$\Theta \models R(b_{i_1}, \ldots, b_{i_{r+1}}, \ldots, b_{i_s})$$

whenever $i_1, \ldots, i_{r+1}$ contain a repeat[13],

$$\Theta \models \neg R(b_{i_1}, \ldots, b_{i_{r+1}}, \ldots, b_{i_s})$$

otherwise. We assume that if L contains any other relation symbols they only occur positively in Θ. Reasoning as before, $\mathcal{S}(\Theta)$ is $\{1, 1, \ldots, 1\}$. Let $\vec{d} \in \{1, 2, 3, \ldots\}^{r+2}$ with no colour appearing more than once. The only s-tuples of colours equivalent according to $\equiv_s^{\bar{E}}$ contain repeats by the definition of $\bar{E}$ since $s \geq r+1$, so for any $i_1, \ldots, i_s$ and $j_1, \ldots, j_s$ from $\{1, \ldots, r+2\}$, if $\langle d_{i_1}, \ldots, d_{i_s}\rangle \equiv_s^{\bar{E}} \langle d_{j_1}, \ldots, d_{j_s}\rangle$, then

$$\Theta \models R(b_{i_1}, \ldots, b_{i_s}) \iff \Theta \models R(b_{j_1}, \ldots, b_{j_s})$$

since they all occur only positively in Θ. So Θ is consistent with (every) such $\vec{d}$ under $\bar{E}$ and $u_{\bar{E}}^{\bar{p},L}(\Theta) > 0$.

On the other hand, let $\Psi(b_1, \ldots, b_r, b_{r+1}, b_{r+2})$ be the state description of L defined by

$$\Psi \models R(b_1, b_{i_2}, \ldots, b_{i_s}) \wedge R(b_{i_1}, b_2, b_{i_3}, \ldots, b_{i_s}) \wedge \cdots \wedge R(b_{i_1}, \ldots, b_{i_r}, b_{r+1}, b_{i_{r+2}}, \ldots, b_{i_s})$$

for all $i_1, \ldots, i_s \in \{1, 2, \ldots, r+2\}$, and

$$\Psi \models \neg R(b_{i_1}, \ldots, b_{i_s})$$

[13] Notice that we are concentrating here just on the initial $r+1$ of the constants instantiating the relation R.

otherwise. We again assume that if L contains other relation symbols they only occur positively in Ψ. Then $\mathcal{S}(\Psi) = \{1, 1, \dots, 1\}$ since

$$\Psi \models R(b_i, b_i, \dots, b_i) \wedge \neg R(b_i, \dots, b_i, b_j, b_i, \dots, b_i)$$

with b_j in the ith place for every $i \in \{1, \dots, r+1\}$, $j \in \{1, \dots, r+2\}$, $i \neq j$, so no two constants can be indistinguishable in Ψ. In addition, for each $\vec{d} \in \{1, 2, 3, \dots\}^{r+2}$, we must have $d_v \equiv_1^{\bar{E}} d_t$ for some $v, t \in \{1, \dots, r+2\}$, $v \neq t$, since there are only $r+1$ equivalence classes under $\equiv_1^{\bar{E}}$. So using (3.14) on the equivalence $\langle d_v, d_v \rangle \equiv_2^{\bar{E}} \langle d_v, d_t \rangle$ gives

$$\langle d_v, d_v, \dots, d_v \rangle \equiv_s^{\bar{E}} \langle d_v, \dots, d_v, d_t, d_v, \dots, d_v \rangle$$

where d_t is in the vth position. If $d_v \equiv_1^{\bar{E}} d_{r+2}$, or if $v \neq t$ are any two from $\{1, \dots, r+1\}$, then (similarly to above)

$$\Psi \models R(b_v, b_v, \dots, b_v) \wedge \neg R(b_v, \dots, b_v, b_t, b_v, \dots, b_v).$$

Therefore, since such v, t exist for every $\vec{d} \in \{1, 2, 3, \dots\}^{r+2}$, Ψ is consistent with no $\vec{d}$ under $\bar{E}$, and $u_{\bar{E}}^{\bar{p},L}(\Psi) = 0$. So $u_{\bar{E}}^{\bar{p},L}(\Theta) \neq u_{\bar{E}}^{\bar{p},L}(\Psi)$ while $\mathcal{S}(\Theta(b_1, \dots, b_{r+2})) = \mathcal{S}(\Psi(b_1, \dots, b_{r+2}))$.

We conclude that the language invariant family $u_{\bar{E}}^{\bar{p},\mathcal{L}}$ satisfies Sx on languages of arity at most r, and no further. $\square$

Chapter 4

PIP and Polyadic Atom Exchangeability

4.1 Introduction

Atom Exchangeability, as introduced on page 15, is a natural symmetry principle in Unary Inductive Logic. It is implied by the Unary Invariance Principle[1] - the symmetry 'umbrella' principle, and implies the previously mentioned Predicate Exchangeability and Strong Negation.

Early proponents of the subject were already aware of it. In his proposed principles of symmetry, Carnap suggested the principle of *Attribute Symmetry* [9, 44]. In broad terms, this is the idea that individuals (constants) are categorised by different families of attributes, and that these should be invariant under permutations of the indexes of each family. Atom Exchangeability can be thought of as the special case when constants are partitioned by the single family 'atoms', and w is invariant under permutations of the names of classes in this family, i.e. the names of atoms.

Further support for this principle was accorded by Carnap and Johnson due to it being a consequence of their favoured Johnson's Sufficientness Postulate [17] (see page 65), and though some criticism of Ax has been raised (for example in [9, 27, 28]), it remains a prominent principle in Unary Inductive Logic.

[1] [36, Proposition 23.5].

In the previous chapter, we introduced the principle of Spectrum Exchangeability as a polyadic generalisation of Atom Exchangeability (under the assumption of Ex). In this chapter, we will show that it is, in fact, PIP that forms the more natural generalisation of Ax. For this purpose, we will introduce the concept of polyadic atoms, a key notion that underpins much of the remainder of this thesis. Results from this chapter appear also in [39] and in [36, Chapter 41].

4.2 Polyadic Atoms

Lemma 2.2 exemplified the unique role state formulae on r variables play in determining automorphisms of BL that permute state formulae for an r-ary language L. We now demonstrate another important role of these formulae; they act as the building blocks of L, much in the same way that atoms act as the building blocks of a unary language. This will allow us to prove the above claim - that PIP is a natural generalisation of the thoroughly studied unary principle of Ax, as stated on page 15 for the language L_q:

Atom Exchangeability, Ax

For any permutation τ of $\{1, 2, \ldots, 2^q\}$ and constants $b_1, \ldots, b_n$,

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\right) = w\left(\bigwedge_{i=1}^{n} \alpha_{\tau(h_i)}(b_i)\right).$$

The formulation of Ax above is the statement that two state descriptions that are mapped one to the other by a permutation of atoms, should get the same probability. As we already mentioned in Chapter 2, it is easy to see that in the unary case, permutations of atoms are in an obvious bijection with permutations of state formulae satisfying (A) and (B) and that *in the unary context* PIP is equivalent to Ax.

We now extend the notion of atoms to polyadic (r-ary) languages L, by defining a *polyadic atom* to be a state formula on r variables. We label the polyadic atoms $\gamma_1(x_1, \ldots, x_r), \gamma_2(x_1, \ldots, x_r), \ldots, \gamma_N(x_1, \ldots, x_r)$ in a fixed order, where the total number of atoms N is $2^{\sum_{d=1}^{q} r^{r_d}}$, since each state formula of r variables contains r^{r_d} conjuncts for each $d = 1, 2, \ldots, q$.

Unless indicated otherwise, γ_k will stand for $\gamma_k(x_1,\dots,x_r)$, with these variables. Note that for purely unary languages, this definition exactly describes the atoms of the language in the original sense (as defined on page 13). We will often drop 'polyadic' and refer to these formulae simply as *atoms*.

In a manner corresponding to the case for unary languages, every state formula of the polyadic language L may be written as a conjunction of polyadic atoms; namely,

$$\Theta(z_1,\dots,z_n) = \bigwedge_{\langle i_1,\dots,i_r\rangle\in\{1,\dots,n\}^r} \gamma_{h_{i_1,\dots,i_r}}(z_{i_1},\dots,z_{i_r}), \tag{4.1}$$

since such a conjunction completely describes the behaviour of[2] $z_1,\dots,z_n$ in relation to each other (and no other variables). In contrast, however, not every such conjunction describes a state formula of L, since some of these will be inconsistent. For instance, for L containing a single binary relation symbol and a state formula $\Theta(z_1,z_2,z_3,z_4)$, we would need $\gamma_{h_{3,4}}(z_3,z_4) = \gamma_{h_{4,3}}(z_4,z_3)$ for the conjunction to be consistent.

Note that when $i_1,\dots,i_r$ in (4.1) are distinct,

$$\gamma_{h_{i_1,\dots,i_r}}(z_{i_1},\dots,z_{i_r}) = \Theta[z_{i_1},\dots,z_{i_r}].$$

On the other hand, when $i_1,\dots,i_r$ are not all distinct we have

$$\gamma_{h_{i_1,\dots,i_r}}(x_1,\dots,x_r) = (\Theta[z_{i_{m_1}},\dots,z_{i_{m_s}}])_\sigma$$

where $i_{m_1},\dots,i_{m_s}$ are the distinct numbers among $i_1,\dots,i_r$, and $\sigma : \{x_1,\dots,x_r\} \twoheadrightarrow \{z_{i_{m_1}},\dots,z_{i_{m_s}}\}$ is defined by $\sigma(x_j) = z_{i_{m_k}} \iff i_j = i_{m_k}$, so

$$\gamma_{h_{i_1,\dots,i_r}}(z_{i_1},\dots,z_{i_r}) = \Theta[z_{i_{m_1}},\dots,z_{i_{m_s}}].$$

4.3 PIP as Polyadic Atom Exchangeability

By Lemma 2.2 every permutation of state formulae that satisfies conditions (A) and (B) from page 20, equivalently a permutation that extends to an automorphism permuting state formulae, is determined by its restriction to the atoms of L. Let Γ denote the

[2]Recall that throughout this account $z_1,\dots,z_n$ denote a distinct choice of variables from $x_1,x_2,\dots$ and that we use '=' also for logical equivalence, as set out on page 11.

set of permutations τ of $\{1, \ldots, N\}$ such that the permutation ξ of atoms defined by $\xi(\gamma_k(x_1, \ldots, x_r)) = \gamma_{\tau(k)}(x_1, \ldots, x_r)$ is a permutation of state formulae satisfying (A) and (B). With these definitions, PIP is clearly equivalent to what may be termed

Polyadic Atom Exchangeability - Permutation Version

For any state description

$$\bigwedge_{\langle i_1,\ldots,i_r\rangle \in \{1,\ldots,n\}^r} \gamma_{h_{i_1,\ldots,i_r}}(b_{i_1}, \ldots, b_{i_r})$$

and $\tau \in \Gamma$,

$$w\left(\bigwedge_{\langle i_1,\ldots,i_r\rangle \in \{1,\ldots,n\}^r} \gamma_{h_{i_1,\ldots,i_r}}(b_{i_1}, \ldots, b_{i_r})\right) = w\left(\bigwedge_{\langle i_1,\ldots,i_r\rangle \in \{1,\ldots,n\}^r} \gamma_{\tau(h_{i_1,\ldots,i_r})}(b_{i_1}, \ldots, b_{i_r})\right).$$

This represents PIP as a generalisation of Ax as stated above, except that we limit the 'allowed' permutations of polyadic atoms to those in Γ. The next result will determine exactly which[3] permutations of atoms define a permutation of state formulae that satisfies (A) and (B).

Lemma 4.1. *A permutation τ of $\{1, \ldots, N\}$ is in Γ if and only if for each $m \leq r$, distinct $1 \leq i_1, \ldots, i_m \leq r$, $\sigma : \{x_1, \ldots, x_r\} \twoheadrightarrow \{x_{i_1}, \ldots, x_{i_m}\}$ and $k, s \in \{1, \ldots, N\}$*

$$(\gamma_k[x_{i_1}, \ldots, x_{i_m}])_\sigma = \gamma_s(x_1, \ldots, x_r) \iff (\gamma_{\tau(k)}[x_{i_1}, \ldots, x_{i_m}])_\sigma = \gamma_{\tau(s)}(x_1, \ldots, x_r). \tag{4.2}$$

Proof. We first show that if τ is in Γ then (4.2) holds. Suppose that $\tau \in \Gamma$ and let η be the associated automorphism of BL. Then $\bar{\eta}$ satisfies (A) and (B) and $\bar{\eta}(\gamma_k) = \gamma_{\tau(k)}$. Assume the left hand side of (4.2) is satisfied. By condition (B)

$$\bar{\eta}(\gamma_k[x_{i_1}, \ldots, x_{i_m}]) = \bar{\eta}(\gamma_k)[x_{i_1}, \ldots, x_{i_m}] = \gamma_{\tau(k)}[x_{i_1}, \ldots, x_{i_m}]$$

so by (A)

$$(\gamma_{\tau(k)}[x_{i_1}, \ldots, x_{i_m}])_\sigma = (\bar{\eta}(\gamma_k[x_{i_1}, \ldots, x_{i_m}]))_\sigma = \bar{\eta}((\gamma_k[x_{i_1}, \ldots, x_{i_m}])_\sigma)\,.$$

[3] Note that the condition in the following lemma is trivial when L is purely unary in accordance with the aforementioned fact that any permutation of unary atoms extends to a permutation of state formulae satisfying (A) and (B).

Therefore, since $\gamma_{\tau(s)} = \bar{\eta}(\gamma_s)$, if $(\gamma_k[x_{i_1}, \ldots, x_{i_m}])_\sigma = \gamma_s(x_1, \ldots, x_r)$ then

$$(\gamma_{\tau(k)}[x_{i_1}, \ldots, x_{i_m}])_\sigma = \bar{\eta}(\gamma_s(x_1, \ldots, x_r)) = \gamma_{\tau(s)}(x_1, \ldots, x_r)$$

so the right hand side of (4.2) holds. Now assume that the right hand side of (4.2) is satisfied. The left hand side follows upon noting that τ^{-1} must also be in Γ since $\bar{\eta}$ has an inverse $\bar{\eta}^{-1}$ which also permutes state formulae [36, Chapter 39], and $\bar{\eta}^{-1}(\gamma_k) = \gamma_{\tau^{-1}(k)}$. So following the above argument with $\bar{\eta}$ replaced by $\bar{\eta}^{-1}$ and γ_k replaced by $\gamma_{\tau(k)}$ yields the required implication. Hence if $\tau \in \Gamma$, the left and right hand sides of (4.2) are equivalent.

To prove the opposite direction, assume that τ satisfies (4.2). We will show that τ is in Γ. First observe that for such τ and for $z_{i_1}, \ldots, z_{i_r}$ *not necessarily distinct* variables, $\gamma_k(z_{i_1}, \ldots, z_{i_r})$ is consistent just when $\gamma_{\tau(k)}(z_{i_1}, \ldots, z_{i_r})$ is consistent. This is the case since for a polyadic atom γ_v, $\gamma_v(z_{i_1}, \ldots, z_{i_r})$ is consistent just when $\gamma_v(x_1, \ldots, x_r)$ is $(\gamma_v[x_{m_1}, \ldots, x_{m_t}])_\sigma$ where $i_{m_1}, \ldots, i_{m_t}$ are the distinct numbers amongst $i_1, \ldots, i_r$ and σ is defined by $\sigma(x_j) = x_{m_u} \iff i_j = i_{m_u}$. Using (4.2), this holds for $\gamma_k(z_{i_1}, \ldots, z_{i_r})$ if and only if it holds for $\gamma_{\tau(k)}(z_{i_1}, \ldots, z_{i_r})$.

Another observation we need is that if two atoms γ_k, γ_h have the property that restricting one to some m variables and the other to some (other) m variables produces the same state formula up to renaming the variables then the same holds for $\gamma_{\tau(k)}, \gamma_{\tau(h)}$. Expressed more formally, for (distinct) $x_{i_1}, \ldots, x_{i_m}$ and $x_{j_1}, \ldots, x_{j_m}$ from $\{x_1, \ldots, x_r\}$ we have

$$\begin{aligned}&\gamma_k[x_{i_1}, \ldots, x_{i_m}] = \gamma_h[x_{j_1}, \ldots, x_{j_m}](x_{i_1}/x_{j_1}, \ldots, x_{i_m}/x_{j_m}) \\ &\qquad \iff \gamma_{\tau(k)}[x_{i_1}, \ldots, x_{i_m}] = \gamma_{\tau(h)}[x_{j_1}, \ldots, x_{j_m}](x_{i_1}/x_{j_1}, \ldots, x_{i_m}/x_{j_m}) \end{aligned} \tag{4.3}$$

where $\gamma_h[x_{j_1}, \ldots, x_{j_m}](x_{i_1}/x_{j_1}, \ldots, x_{i_m}/x_{j_m})$ is the result of replacing every occurrence in $\gamma_h[x_{j_1}, \ldots, x_{j_m}]$ of x_{j_v} by x_{i_v}, $v = 1, \ldots, m$.

To see this, consider for example $\sigma_1 : \{x_1, \ldots, x_r\} \twoheadrightarrow \{x_{i_1}, \ldots, x_{i_m}\}$ and $\sigma_2 : \{x_1, \ldots, x_r\} \twoheadrightarrow \{x_{j_1}, \ldots, x_{j_m}\}$ defined by

$$\sigma_1(x_i) = \begin{cases} x_i & \text{if } i \in \{i_1, \ldots, i_m\}, \\ x_{i_1} & \text{otherwise,} \end{cases} \qquad \sigma_2(x_i) = \begin{cases} x_{j_v} & \text{if } i = i_v \in \{i_1, \ldots, i_m\}, \\ x_{j_1} & \text{otherwise.} \end{cases}$$

Then the left hand side of (4.3) holds just if for these σ_1, σ_2 we have

$$(\gamma_k[x_{i_1}, \ldots, x_{i_m}])_{\sigma_1} = \gamma_s(x_1, \ldots, x_r) = (\gamma_h[x_{j_1}, \ldots, x_{j_m}])_{\sigma_2}$$

for some $\gamma_s(x_1, \ldots, x_r)$, in which case

$$(\gamma_{\tau(k)}[x_{i_1}, \ldots, x_{i_m}])_{\sigma_1} = \gamma_{\tau(s)}(x_1, \ldots, x_r) = (\gamma_{\tau(h)}[x_{j_1}, \ldots, x_{j_m}])_{\sigma_2}$$

follows by (4.2), implying the right hand side of the equivalence. The other direction follows similarly.

For a state formula

$$\Theta(z_1, \ldots, z_n) = \bigwedge_{\langle i_1,\ldots,i_r\rangle \in \{1,\ldots,n\}^r} \gamma_{h_{i_1,\ldots,i_r}}(z_{i_1}, \ldots, z_{i_r}),$$

let the function $\digamma$ be defined by

$$\digamma(\Theta(z_1, \ldots, z_n)) = \bigwedge_{\langle i_1,\ldots,i_r\rangle \in \{1,\ldots,n\}^r} \gamma_{\tau(h_{i_1,\ldots,i_r})}(z_{i_1}, \ldots, z_{i_r}). \tag{4.4}$$

By the first of the above observations each conjunct in (4.4) is consistent. Moreover, the whole conjunction must be consistent, since otherwise there would be $\langle i_1, \ldots, i_r\rangle$ and $\langle j_1, \ldots, j_r\rangle$ from $\{1, \ldots, n\}^r$ and distinct $k_1, \ldots, k_t$ occurring both amongst $\{i_1, \ldots, i_r\}$ and $\{j_1, \ldots, j_r\}$ such that for some relation symbol R_d of L of arity r_d and some $m_1, \ldots, m_{r_d}$ from $\{1, \ldots, t\}$,

$$\gamma_{\tau(h_{i_1,\ldots,i_r})}(z_{i_1}, \ldots, z_{i_r}) \models R_d(z_{k_{m_1}}, \ldots, z_{k_{m_{r_d}}}),$$
$$\gamma_{\tau(h_{j_1,\ldots,j_r})}(z_{j_1}, \ldots, z_{j_r}) \models \neg R_d(z_{k_{m_1}}, \ldots, z_{k_{m_{r_d}}}).$$

This would mean that

$$\gamma_{\tau(h_{i_1,\ldots,i_r})}(z_{i_1}, \ldots, z_{i_r})[z_{k_1}, \ldots, z_{k_t}] \neq \gamma_{\tau(h_{j_1,\ldots,j_r})}(z_{j_1}, \ldots, z_{j_r})[z_{k_1}, \ldots, z_{k_t}],$$

so by the second observation

$$\gamma_{h_{i_1,\ldots,i_r}}(z_{i_1}, \ldots, z_{i_r})[z_{k_1}, \ldots, z_{k_t}] \neq \gamma_{h_{j_1,\ldots,j_r}}(z_{j_1}, \ldots, z_{j_r})[z_{k_1}, \ldots, z_{k_t}].$$

However this is impossible since both are $\Theta(z_1, \ldots, z_n)[z_{k_1}, \ldots, z_{k_t}]$. Therefore $\digamma$ defined by (4.4) permutes state formulae and clearly $\digamma(\gamma_k(x_1, \ldots, x_r)) = \gamma_{\tau(k)}(x_1, \ldots, x_r)$.

It remains to check that $\digamma$ satisfies conditions (A) and (B). Condition (B) clearly holds and for (A), let

$$\Theta(z_1, \dots, z_m) = \bigwedge_{\langle i_1,\dots,i_r\rangle \in \{1,\dots,m\}^r} \gamma_{h_{i_1,\dots,i_r}}(z_{i_1}, \dots, z_{i_r})$$

and let $\sigma : \{y_1, \dots, y_n\} \twoheadrightarrow \{z_1, \dots, z_m\}$. Writing σ also for the mapping from $\{1, \dots, n\}$ to $\{1, \dots, m\}$ that sends j to i iff $\sigma(y_j) = z_i$, we have

$$(\Theta(z_1, \dots, z_m))_\sigma = \bigwedge_{\langle j_1,\dots,j_r\rangle \in \{1,\dots,n\}^r} \gamma_{h_{\sigma(j_1),\dots,\sigma(j_r)}}(y_{j_1}, \dots, y_{j_r}),$$

since Θ_σ is the (unique) state formula Φ such that $\Phi(\sigma(y_1), \dots, \sigma(y_n)) = \Theta(z_1, \dots, z_m)$ and $\gamma_{h_{\sigma(j_1),\dots,\sigma(j_r)}}(\sigma(y_{j_1}), \dots, \sigma(y_{j_r})) = \gamma_{h_{i_1,\dots,i_r}}(z_{i_1}, \dots, z_{i_r})$. In addition,

$$\digamma(\Theta(z_1, \dots, z_m)) = \bigwedge_{\langle i_1,\dots,i_r\rangle \in \{1,\dots,m\}^r} \gamma_{\tau(h_{i_1,\dots,i_r})}(z_{i_1}, \dots, z_{i_r}).$$

So both $(\digamma(\Theta(z_1, \dots, z_m)))_\sigma$ and $\digamma((\Theta(z_1, \dots, z_m))_\sigma)$ are

$$\bigwedge_{\langle j_1,\dots,j_r\rangle \in \{1,\dots,n\}^r} \gamma_{\tau(h_{\sigma(j_1),\dots,\sigma(j_r)})}(y_{j_1}, \dots, y_{j_r})$$

and thus (A) holds. □

We now show that another formulation of Ax, which in the unary case is easily seen to be equivalent to the one given above, in the polyadic context becomes a principle that is *not obviously* equivalent to PIP but somewhat surprisingly turns out to be so nevertheless.

Atom Exchangeability (II)

Let

$$\Theta(b_1, \dots, b_n) = \bigwedge_{i=1}^{n} \alpha_{h_i}(b_i), \qquad \Phi(b_1, \dots, b_n) = \bigwedge_{i=1}^{n} \alpha_{k_i}(b_i)$$

be state descriptions of a unary language. If for all $0 \leq i, j \leq n$ *we have*

$$h_i = h_j \Longleftrightarrow k_i = k_j$$

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

The immediate polyadic counterpart of this is

Polyadic Atom Exchangeability - Spectral-Equivalence Version, PAx

Let

$$\Theta(b_1, \dots, b_n) = \bigwedge_{\langle i_1,\dots,i_r\rangle \in \{1,\dots,n\}^r} \gamma_{h_{i_1,\dots,i_r}}(b_{i_1}, \dots, b_{i_r})$$

and

$$\Phi(b_1, \dots, b_n) = \bigwedge_{\langle i_1,\dots,i_r\rangle \in \{1,\dots,n\}^r} \gamma_{k_{i_1,\dots,i_r}}(b_{i_1}, \dots, b_{i_r})$$

be state descriptions of L such that for all $\langle i_1, \dots, i_r\rangle, \langle j_1, \dots, j_r\rangle \in \{1, \dots, n\}^r$

$$h_{i_1,\dots,i_r} = h_{j_1,\dots,j_r} \iff k_{i_1,\dots,i_r} = k_{j_1,\dots,j_r}. \tag{4.5}$$

Then $w(\Theta(b_1, \dots, b_n)) = w(\Phi(b_1, \dots, b_n))$.

We shall show that PIP is equivalent to PAx and in order to do so we will use the results on similarity from Section 2.4.

Theorem 4.2. *The principle PAx is equivalent to PIP.*

Proof. First assume that w satisfies PAx. Suppose that $\digamma$ is a permutation of state formulae that satisfies (A) and (B), Θ is a state formula and $\Phi = \digamma(\Theta)$. Assuming $\Theta(b_1, \dots, b_n)$ and $\Phi(b_1, \dots, b_n)$ are written as in the statement of PAx, by condition (B) we have that $\digamma(\gamma_{h_{i_1,\dots,i_r}}) = \gamma_{k_{i_1,\dots,i_r}}$ so (4.5) holds. Hence

$$w(\Theta(b_1, \dots, b_n)) = w(\Phi(b_1, \dots, b_n)),$$

showing PIP for w.

Now suppose that w satisfies PIP. Let Θ, Φ be as in the statement of PAx and such that (4.5) holds. It suffices to show that Θ and Φ are similar since then it will follow by Theorem 2.6 and PIP that $w(\Theta(b_1, \dots, b_n)) = w(\Phi(b_1, \dots, b_n))$, proving PAx for w.

So suppose that for distinct $i_1, \dots, i_t$ and $j_1, \dots, j_s$ from $\{1, \dots, n\}$ and $\sigma : \{z_{i_1}, \dots, z_{i_t}\} \twoheadrightarrow \{z_{j_1}, \dots, z_{j_s}\}$ we have

$$\Theta[z_{i_1}, \dots, z_{i_t}] = (\Theta[z_{j_1}, \dots, z_{j_s}])_\sigma.$$

Then for every choice of $m_1, \dots, m_{r_d}$ (with possible repeats) from $\{i_1, \dots, i_t\}$ and each relation symbol R_d of arity r_d,

$$\Theta \models R_d(z_{m_1}, \dots, z_{m_{r_d}}) \iff \Theta \models R_d(\sigma(z_{m_1}), \dots, \sigma(z_{m_{r_d}}))$$

since

$$\Theta[z_{i_1},\dots,z_{i_t}] \models R_d(z_{m_1},\dots,z_{m_{r_d}}) \Longleftrightarrow \Theta[z_{j_1},\dots,z_{j_s}] \models R_d(\sigma(z_{m_1}),\dots,\sigma(z_{m_{r_d}})).$$

With a slight abuse of notation, writing $\sigma(i_k) = j_e$ instead of $\sigma(z_{i_k}) = z_{j_e}$, this means that for any $m_1,\dots,m_r$ (with possible repeats) from $\{i_1,\dots,i_t\}$ we have $h_{m_1,\dots,m_r} = h_{\sigma(m_1),\dots,\sigma(m_r)}$, as $\gamma_{h_{m_1,\dots,m_r}}$ describes every relation involving variables from $\{z_{m_1},\dots,z_{m_r}\}$ and similarly for $\gamma_{h_{\sigma(m_1),\dots,\sigma(m_r)}}$.

If we had

$$\Phi[z_{i_1},\dots,z_{i_t}] \neq (\Phi[z_{j_1},\dots,z_{j_s}])_\sigma$$

then reasoning as above, this would mean that for some $m_1,\dots,m_r$ from $\{i_1,\dots,i_t\}$, $k_{m_1,\dots,m_r} \neq k_{\sigma(m_1),\dots,\sigma(m_r)}$. However this would contradict (4.5), so $\Phi[z_{i_1},\dots,z_{i_t}] = (\Phi[z_{j_1},\dots,z_{j_s}])_\sigma$ and since the same argument can be repeated with Θ and Φ interchanged, we conclude that Θ and Φ are similar as required. □

4.4 PIP as Polyadic Atom Exchangeability *under Ex*

We have thus far shown that two versions of Atom Exchangeability on unary languages result in the principle PIP on polyadic languages when formulated using polyadic atoms. The third remaining formulation of Ax in the unary context utilises the idea of a *spectrum* of a state description, as explained on page 34 and restated below. This version can easily be seen to be equivalent to the previous statements of unary Ax *if we assume that Ex holds.* It would be natural to ask therefore, whether a polyadic formulation of this version of Ax would be equivalent to PIP + Ex. We shall show that for the most immediate polyadic counterpart of this principle the answer would be no. Whether other possible definitions of polyadic spectrum do indeed provide an equivalence with PIP + Ex remains a topic for further research.

Atom Exchangeability (III)

For $\Theta(b_1, \dots, b_n)$ a state description of a unary language L_q, the probability

$$w(\Theta(b_1, \dots, b_n)) = w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\right)$$

depends only on the spectrum *of this state description, that is on the multiset* $\{m_1, \dots, m_{2^q}\}$ *where* $m_j = |\{i \mid h_i = j\}|$.

By analogy, in the polyadic case this gives rise to defining the *p-spectrum* (polyadic, atom-based spectrum) of a state description

$$\Theta(b_1, \dots, b_n) = \bigwedge_{\langle i_1, \dots, i_r\rangle \in \{1,\dots,n\}^r} \gamma_{h_{i_1,\dots,i_r}}(b_{i_1}, \dots, b_{i_r})$$

of a polyadic language L as the multiset $\{m_1, \dots, m_N\}$ where

$$m_j = |\{\langle i_1, \dots, i_r\rangle \in \{1, \dots, n\}^r \mid h_{i_1,\dots,i_r} = j\}|.$$

For ease of notation, we usually omit zero entries from our multisets.

We remark that current use of the term *spectrum* in Polyadic Inductive Logic, as in the statement of Spectrum Exchangeability (on page 34), which involves the strong notion of indistinguishability of constants in a particular state description, is clearly different from the notion of p-spectrum. Unless the language is unary, this type of indistinguishability is *not preserved* when the state description is extended, that is when we consider another state description with additional constants that implies the given one.

On the other hand, in the definition of a p-spectrum of a state description we consider ordered r-tuples of constants (possibly with repeats), classifying them purely by the way these r constants relate to each other in the state description and disregarding their connections to the other constants. If we choose to define p-indistinguishability of two r-tuples in a state description to mean satisfying the same atom within it, we find that this notion of p-indistinguishability is 'forever': extending the state description to more constants does not change it.

With this in mind, we arrive at the following new polyadic symmetry principle:

Atom-based Spectrum Exchangeability, p-Sx

The probability of a state description of a polyadic language L depends only on its p-spectrum.

Examining this new principle, we can see that p-Sx implies Ex, since if w satisfies p-Sx then $w(\Theta(b_1, \dots, b_n)) = w(\Theta(b'_1, \dots, b'_n))$ when $b_1, \dots, b_n$ and $b'_1, \dots, b'_n$ are distinct constants and $\Theta(b_1, \dots, b_n)$ is a state description, because they share the same p-spectrum. It also implies PAx (and hence PIP), since any two state descriptions that satisfy (4.5) necessarily have the same p-spectrum.

We now show the converse does not hold in general by pointing out a probability function that satisfies PIP + Ex but gives different probabilities to state descriptions with the same p-spectrum. For this purpose, we employ one of the probability functions $u^{\bar{p},L}_{\bar{E}}$ used in the previous chapter.

As in the proof of Proposition 3.5, let L be a language with a single binary relation symbol R. Let $\bar{p} = \left\langle 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, \dots \right\rangle$ and let $\bar{E}$ contain the equivalence $\langle 1, 2 \rangle \equiv^{\bar{E}}_2 \langle 3, 4 \rangle$ together with all equivalences that follow by condition (3.4).

Having picked the sequence $\langle c_1, \dots, c_n \rangle$ where each c_j is chosen to be i with probability p_i, we pick uniformly at random a state description consistent with this sequence under $\bar{E}$, where a state description

$$\Theta(b_1, \dots, b_n) \;=\; \bigwedge_{\langle i_1, i_2 \rangle \in \{1, \dots, n\}^2} \gamma_{h_{i_1, i_2}}(b_{i_1}, b_{i_2})$$

of L is *consistent with* $\langle c_1, \dots, c_n \rangle$ *under* $\bar{E}$ if[4] for any $\langle i_1, i_2 \rangle, \langle j_1, j_2 \rangle \in \{1, \dots, n\}^2$,

$$\langle c_{i_1}, c_{i_2} \rangle \equiv^{\bar{E}}_2 \langle c_{j_1}, c_{j_2} \rangle \implies h_{i_1, i_2} = h_{j_1, j_2}. \tag{4.6}$$

Then (as before), $u^{\bar{p},L}_{\bar{E}}(\Theta(b_1, \dots, b_n))$ is the probability $\Theta(b_1, \dots, b_n)$ is the state description arrived at by the above process.

[4]This definition for an r-ary language is equivalent to Definition 3.4 given on page 38.

Let $\Theta(b_1, b_2, b_3, b_4)$ and $\Phi(b_1, b_2, b_3, b_4)$ be state descriptions represented respectively by the following matrices

$$\begin{matrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{matrix} \qquad\qquad \begin{matrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{matrix}.$$

The p-spectrum of both is $\{10, 6\}$, so it remains to show that $u^{\bar{p},L}_{\bar{E}}(\Theta) \neq u^{\bar{p},L}_{\bar{E}}(\Phi)$.

To see this, note that neither Θ nor Φ are consistent with any sequence of colours $\langle c_1, c_2, c_3, c_4 \rangle$ in which a colour appears more than once, since as explained by the remark on page 42, this would require the state descriptions to have indistinguishable constants and that is not the case.

So consider a sequence $\langle c_1, c_2, c_3, c_4 \rangle$ where each colour appears exactly once. For some permutation ν of $\{1, 2, 3\}$ we must have $\langle c_{\nu(1)}, c_{\nu(2)} \rangle \equiv^{\bar{E}}_2 \langle c_{\nu(3)}, c_4 \rangle$ but

$$\Theta[b_{\nu(1)}, b_{\nu(2)}] = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}, \qquad \Theta[b_{\nu(3)}, b_4] = \begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}$$

for *every* ν, so Θ is consistent with no sequence $\langle c_1, c_2, c_3, c_4 \rangle$ and hence $u^{\bar{p},L}_{\bar{E}}(\Theta) = 0$.

On the other hand, Φ is consistent for example with the sequence $\langle 1, 2, 3, 4 \rangle$ and hence $u^{\bar{p},L}_{\bar{E}}(\Phi) \neq 0$.[5] Thus $u^{\bar{p},L}_{\bar{E}}$ is a function that satisfies PIP and Ex without satisfying p-Sx, as claimed.

[5]The total probabilities may be checked to be $u^{\bar{p},L}_{\bar{E}}(\Theta) = 0$ and $u^{\bar{p},L}_{\bar{E}}(\Phi) = 2^{-16}$.

Chapter 5

Binary Signature Exchangeability

5.1 Introduction

We investigate the notion of a *signature* in binary Inductive Logic, introduce the *Principle of Signature Exchangeability* and study the probability functions satisfying it. We prove a representation theorem for such functions and show that they satisfy a binary version of the Principle of Instantial Relevance. In the next chapter, we extend this investigation to general polyadic languages. The material in this chapter appears also in [40].

We begin with a closer inspection of the principle of Constant Exchangeability, which we have already met. Known by Johnson as *The Permutation Postulate* (see for example [17, 43, 44]) or in Carnap's terms, *the Principle of Symmetry* [6, 8], Ex is a widely accepted and commonly assumed rational requirement in Pure Inductive Logic. Informally, this is the statement that in the absence of further knowledge, different individuals of our universe should be treated equally. In our framework it means that the probability assigned to a sentence is independent of the particular constants instantiating it. In addition, in the thoroughly studied unary context, this principle exists in an equivalent formulation - as invariance under signatures of state descriptions. This unary characterisation of the principle has led to some of the most significant results in Unary Inductive Logic thus far. These include, for example, a complete characterisation of functions satisfying Ex, and the Principle of Instantial Relevance (see page 64) following as a logical consequence of Constant Exchangeability.

In contrast, such results have so far not translated satisfactorily into the polyadic. Having extended the concept of atoms to polyadic languages in the previous chapter, in the following chapters we generalise the notion of a signature first to binary and then to Polyadic Inductive Logic, and investigate the theory this yields for higher arity languages. As has been our custom, in this first section we give a brief account of the relevant background which in this case is the unary portion we shall be concerned with for the purpose of this chapter. We then suggest new methods and formulations for these concepts for binary languages. Specifically, we present a binary definition of a signature and a principle of invariance under this notion, an independence principle characterising the basic functions satisfying this new signature-based principle, and a binary version of the Principle of Instantial Relevance.

Recall the Principle of Constant Exchangeability, as stated on page 14, where we also remarked that it can be equivalently expressed as requiring (1.2) to hold only for state descriptions Θ instead of general $\theta \in SL$. This leads to a simpler formulation of Ex for unary languages (as mentioned above), based on the notion of a *signature*[1]. The signature of a state description $\Theta(b_1, \ldots, b_m) = \bigwedge_{i=1}^m \alpha_{h_i}(b_i)$ is defined to be the vector $\langle m_1, \ldots, m_{2^q} \rangle$ where m_j is the number of times that α_j appears amongst the α_{h_i}. Ex in the unary case thus amounts to[2]

Constant Exchangeability, unary version

The probability of a state description depends only on its signature.

In the previous chapter, we investigated atom-based polyadic approaches to the principle of Atom Exchangeability. We now mention a collection of other important principles from Unary Inductive Logic that are stated in terms of (unary) atoms.

Principle of Instantial Relevance, PIR

$$w\left(\alpha_j(a_{m+2}) \mid \bigwedge_{i=1}^m \alpha_{h_i}(a_i)\right) \leq w\left(\alpha_j(a_{m+2}) \mid \alpha_j(a_{m+1}) \wedge \bigwedge_{i=1}^m \alpha_{h_i}(a_i)\right).$$

This principle was suggested by Carnap [7, Chapter 13] and expresses the idea that having witnessed an event in the past should enhance (or at least should not decrease)

[1] see [36, Chapter 8].

[2] Ex implies that the probability of a unary state description depends only on the atoms occurring in it, and not which constants instantiate them.

our belief that we might see it again in future.

Johnson's Sufficientness Postulate, JSP

$w(\alpha_j(a_{m+1}) \mid \bigwedge_{i=1}^m \alpha_{h_i}(a_i))$ *depends only on* m *and on* m_j*, where* m_j *is the number of times that* α_j *appears amongst the* α_{h_i}*.*

First appearing in [17], JSP states that our belief in seeing an individual with a certain combination of properties should depend only on how many individuals we have seen, and how many of them have satisfied exactly the same combination of properties.

Unary Principle of Induction, UPI

Assume that $m_j \leq m_s$*, where* m_j*,* m_s *are the numbers of times that* α_j*,* α_s *respectively appear amongst the* α_{h_i}*. Then*

$$w\left(\alpha_j(a_{m+1}) \mid \bigwedge_{i=1}^m \alpha_{h_i}(a_i)\right) \leq w\left(\alpha_s(a_{m+1}) \mid \bigwedge_{i=1}^m \alpha_{h_i}(a_i)\right).$$

This principle [36, Chapter 21] says that if we have already seen at least as many individuals with a certain combination of properties as with another combination, we should think the next individual at least as likely to have the first combination of properties as the second.

Finally, we mention the (not necessarily unary) Constant Irrelevance or Independence Principle. It is not stated in terms of atoms, but it plays a role in what follows.

Constant Independence Principle, IP

Let $\theta, \phi \in QFSL$ *have no constant symbols in common. Then*

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

In the unary context [36, Chapter 8], the only probability functions satisfying IP together with Ex are the $w_{\vec{x}}$ functions, where $\vec{x} = \langle x_1, \ldots, x_{2^q} \rangle$ is from

$$\mathbb{D}_{2^q} \;=\; \left\{ \langle x_1, \ldots, x_{2^q} \rangle \mid x_1, x_2, \ldots, x_{2^q} \geq 0 \text{ and } \sum_{j=1}^{2^q} x_j = 1 \right\}$$

and $w_{\vec{x}}$ is determined by

$$w_{\vec{x}}\left(\bigwedge_{i=1}^m \alpha_{h_i}(b_i)\right) \;=\; \prod_{i=1}^m x_{h_i} \;=\; \prod_{j=1}^{2^q} x_j^{m_j},$$

where m_j is again the number of times that α_j appears amongst the α_{h_i}. Thus $w_{\vec{x}}$ is the (unique) function that assigns the probability x_j to all $\alpha_j(a_i)$ regardless of i, and treats instantiations of atoms (both the same or different) by distinct constants as stochastically independent. These functions are remarkably useful because they are simple and since all unary probability functions satisfying Ex can be generated from them as continuous convex combinations (integrals). The precise statement of this claim [12] is

de Finetti's Representation Theorem. *Let L be a unary language with q predicate symbols and let w be a probability function on SL satisfying Ex. Then there is a normalised, σ-additive measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ such that*

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

for any state description Θ of L, and conversely, given such a μ, w as above extends uniquely to a probability function on SL satisfying Ex.

Early results of Unary Inductive Logic show that any probability function satisfying Ex also satisfies PIR (as already mentioned, [13]), and that - provided the language has at least two predicate symbols - any probability function satisfying Ex and JSP must be one of the Carnap Continuum functions[3]. A later result due to Paris and Waterhouse [37] shows that any probability function satisfying Ex and Ax must also satisfy UPI.

These are pleasing results in Pure Inductive Logic, since we know that if we make these rational requirements, we also gain their consequences - a PIL version of 'buy one (or two), get one free'. So, for example, if we are happy to accept Ex and Ax we also gain the appealing UPI.

5.2 An Atom-based Approach for Binary Languages

We shall now consider how atoms can aid us to understand the properties of probability functions on binary languages. Let L contain some binary relation symbols and possibly some unary predicate symbols, but no symbols of higher arity. We shall denote

[3]References for these functions were mentioned on page 17.

the unary predicate symbols by $P_1, \ldots, P_{q_1}$ and the binary symbols by $Q_1, \ldots, Q_{q_2}$, with $q_1 + q_2 = q$.

In this language, the state formulae *for one variable* have the form

$$\bigwedge_{i=1}^{q_1} \pm P_i(x) \wedge \bigwedge_{u=1}^{q_2} \pm Q_u(x,x) \tag{5.1}$$

and we will write

$$\beta_1(x), \ldots, \beta_{2^q}(x)$$

for them (using the usual lexicographic ordering). We also refer to these formulae as *1-atoms* (since they act on one individual). There are 2^q many of these since there are q relation symbols and each of them can appear in β_k either positively or negatively.

The *atoms* of the language, that is, the state formulae for two variables, have the form

$$\beta_k(x) \wedge \beta_c(y) \wedge \bigwedge_{u=1}^{q_2} \pm Q_u(x,y) \wedge \bigwedge_{u=1}^{q_2} \pm Q_u(y,x).$$

There are $N = 2^{2q}2^{2q_2}$ atoms, and we shall denote them by

$$\gamma_1(x,y), \ldots, \gamma_N(x,y)\,.$$

In order to help visualise the binary case, we introduce the notation $\delta_s(x,y)$ for the conjunctions $\bigwedge_{u=1}^{q_2} \pm Q_u(x,y)$, where $s = 1, \ldots, 2^{q_2}$ (and the δ_s are again ordered lexicographically). Any atom $\gamma_h(x,y)$ can then be written as

$$\beta_k(x) \wedge \beta_c(y) \wedge \delta_e(x,y) \wedge \delta_d(y,x) \tag{5.2}$$

for some $1 \leq k, c \leq 2^q$, $1 \leq e, d \leq 2^{q_2}$. We shall represent such an atom by the matrix

$$\begin{pmatrix} k & e \\ d & c \end{pmatrix},$$

and write $\gamma_{[k,c,e,d]}(x,y)$ for this atom (5.2). We refer to $\beta_k(x) \wedge \beta_c(y)$ as the *unary trace* of the atom (5.2).

In addition, we assume that the atoms of L are ordered unambiguously: when k, c run through $1, \ldots, 2^q$ and e, d run through $1, \ldots, 2^{q_2}$, the number

$$2^{q+2q_2}(k-1) + 2^{2q_2}(c-1) + 2^{q_2}(e-1) + d \tag{5.3}$$

runs through $1, \dots, 2^{2q}2^{2q_2}$. Then there is exactly one way of obtaining each of the numbers $1, \dots, N$, that is one value of k, c, e, d which gives each of the atoms.

Example. When L has just one, binary, relation symbol Q (that is, when $q_1 = 0$, $q_2 = 1$) then $\beta_1(x)$ and $\beta_2(x)$ are $Q(x,x)$ and $\neg Q(x,x)$ respectively, and $\delta_1(x,y)$ and $\delta_2(x,y)$ are $Q(x,y)$ and $\neg Q(x,y)$ respectively. One possible atom of this language is

$$Q(x,x) \wedge Q(y,y) \wedge \neg Q(x,y) \wedge Q(y,x),$$

which corresponds to the atom $\gamma_3(x,y)$ and it is represented by the matrix

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

In this very special case of a language containing just one binary relation symbol, we often write 0 in place of 2 (hence Q and $\neg Q$ correspond to 1 and 0 respectively, as we had previously), so the atom above may also be represented by

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Using atoms, a state description of L can be written as

$$\Theta(b_1, \dots, b_m) = \bigwedge_{1 \le i,t \le m} \gamma_{h_{i,t}}(b_i, b_t) \tag{5.4}$$

and it can be represented by an $m \times m$ matrix

$$\begin{pmatrix} k_1 & e_{1,2} & e_{1,3} & \dots & e_{1,m} \\ d_{1,2} & k_2 & e_{2,3} & \dots & e_{2,m} \\ d_{1,3} & d_{2,3} & k_3 & \dots & e_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1,m} & d_{2,m} & d_{3,m} & \dots & k_m \end{pmatrix} \tag{5.5}$$

for some

$$1 \le k_i \le 2^q, \quad 1 \le e_{i,t}\,,\, d_{i,t} \le 2^{q_2}.$$

This means that depending on whether $i < t$ or $t < i$, $\gamma_{h_{i,t}}$ is

$$\begin{pmatrix} k_i & e_{i,t} \\ d_{i,t} & k_t \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} k_t & d_{i,t} \\ e_{i,t} & k_i \end{pmatrix}$$

respectively, and $\gamma_{h_{i,i}}$ is

$$\begin{pmatrix} k_i & e \\ e & k_i \end{pmatrix}$$

for that e for which $\Theta(b_1, \ldots, b_m) \models \delta_e(b_i, b_i)$. Notice that this is the case since $\gamma_{h_{i,i}}$ is equivalent to some β_k as in (5.1) and there are indeed $2^{q_1}2^{q_2} = 2^q$ choices of these.

Clearly, there is much over-specification in the expression (5.4); for example, we must have $\gamma_{h_{t,i}}(x, y) = \gamma_{h_{i,t}}(y, x)$. A more efficient way of writing a state description (for at least two individuals) in terms of atoms is to restrict i, t in (5.4) to $i < t$,

$$\Theta(b_1, \ldots, b_m) = \bigwedge_{1 \leq i < t \leq m} \gamma_{h_{i,t}}(b_i, b_t). \tag{5.6}$$

This contains all the information about Θ and it still over-specifies all that concerns single individuals. In this investigation we will find it convenient to make this part of the state description visible, so we shall write it as

$$\Theta(b_1, \ldots, b_m) = \bigwedge_{1 \leq i \leq m} \beta_{k_i}(b_i) \wedge \bigwedge_{1 \leq i < t \leq m} \gamma_{h_{i,t}}(b_i, b_t). \tag{5.7}$$

This works even when $m = 1$. We adopt a convention that if needed we still write $\gamma_{h_{t,i}}(x, y)$ for $\gamma_{h_{i,t}}(y, x)$.

Definition 5.1. For Θ as in (5.7), we define

$$\bigwedge_{1 \leq i \leq m} \beta_{k_i}(b_i) \tag{5.8}$$

to be the *unary trace* of Θ. Any conjunction of this form is called a *unary trace for* $b_1, \ldots, b_m$.

We remark that when using atoms, some over-specification is unavoidable. It is possible to develop an approach to Polyadic Inductive Logic using just *elements* rather than atoms (where elements in the binary case are the conjunctions $\bigwedge_{i=1}^{q_1} \pm P_i(x)$ and the conjunctions $\bigwedge_{u=1}^{q_2} \pm Q_u(x, y)$ - these are the $\delta_s(x, y)$, and analogously for higher arity languages), and thus to avoid over-specification. We will do this in Chapter 7, where we will also mention some of the advantages and disadvantages of this approach. In short however, such a 'disjointed' approach fails to capture much of the structure of the sentences we wish to work with. For example, in the disjointed approach, the ordered pairs obtained from each other by changing the order of the two individuals are treated separately, and consequently some crucial connections are lost.

In order to develop a binary approach to the principles we mentioned in Section 5.1 we will need also the concept of a *partial state description*. These are sentences which, like state descriptions, specify all that can be said about all single individuals from amongst the $b_1, \ldots, b_m$, and all that can be said about *some* pairs of them:

Definition 5.2. A *partial state description for* $b_1, \ldots, b_m$ *is a sentence*

$$\Delta(b_1, \ldots, b_m) = \bigwedge_{1 \leq i \leq m} \beta_{k_i}(b_i) \wedge \bigwedge_{\substack{\{b_i, b_t\} \in C \\ i<t}} \gamma_{h_{i,t}}(b_i, b_t)\,, \tag{5.9}$$

where C is some set of 2-element subsets of $\{b_1, \ldots, b_m\}$.

We use capital Greek letters also for partial state descriptions.

Example. Using the representation described above for L containing just one binary relation symbol Q, the matrix

$$\begin{array}{cccc} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}$$

represents the (full) state description

$$\bigwedge_{i,t=1}^{3} Q(b_i, b_t) \wedge \bigwedge_{i=1}^{3} (\neg Q(b_i, b_4) \wedge \neg Q(b_4, b_i)) \wedge Q(b_4, b_4)\,,$$

while

$$\begin{array}{cccc} 1 & & 1 & \\ & 1 & & \\ 0 & & 1 & 0 \\ & & 0 & 1 \end{array}$$

represents the partial state description

$$\bigwedge_{i=1}^{4} Q(b_i, b_i) \wedge (Q(b_1, b_3) \wedge \neg Q(b_3, b_1)) \wedge (\neg Q(b_3, b_4) \wedge \neg Q(b_4, b_3)).$$

The matrix

$$\begin{array}{cccc} 1 & & 1 & \\ & 1 & 0 & \\ 0 & & 1 & 0 \\ & & 0 & 1 \end{array}$$

represents no partial state description since it gives some - but not all - information about the pair b_2, b_3*. Specifying also* $Q(b_3, b_2)$ *or* $\neg Q(b_3, b_2)$ *would turn it into a partial state description.*

We remark that if C in (5.9) contains no 2-element subsets, that is $C = \emptyset$, then (5.9) is still a partial state description. In particular, a unary trace for $b_1, \ldots, b_m$ is a partial state description for $b_1, \ldots, b_m$. Of course, every state description is also a partial state description. Secondly, we mention that *partial state formulae* are defined analogously to partial state descriptions, with $b_1, \ldots, b_m$ replaced by (distinct) variables $z_1, \ldots, z_m$. Finally, we follow the convention that only the individuals that are mentioned after some β_{k_i} in (5.9) are listed in brackets after Θ, and that they are distinct.

5.3 Binary Signatures

In Unary Inductive Logic, it is almost always the case that Ex is assumed. If we wish to continue assuming Ex and to base our theory on polyadic atoms, we need to be able to work with the atoms in a way which reflects that atoms obtained from each other by permuting the variables are in some sense equivalent and represent the same thing.

In the binary case, atoms have two variables and there is only one non-trivial permutation of $\{x, y\}$. If $\gamma(x, y)$ is the atom represented by

$$\begin{pmatrix} k & e \\ d & c \end{pmatrix}$$

then permuting x and y yields the atom represented by

$$\begin{pmatrix} c & d \\ e & k \end{pmatrix}.$$

If $k = c$ and $e = d$ then these are the same atom.

Hence, when wishing to disregard the order, the behaviour of pairs of individuals should be classified by the atom they satisfy, only up to the equivalence defined on atoms by

$$\begin{pmatrix} k & e \\ d & c \end{pmatrix} \sim \begin{pmatrix} c & d \\ e & k \end{pmatrix}.$$

That is, $\gamma_{[k,c,e,d]} \sim \gamma_{[c,k,d,e]}$.

This means that rather than N different ways a pair can behave, there are $p < N$ of them, where N is the number of atoms and p is the number of $\sim$-equivalence classes. Explicitly, $p = (N + 2^q \cdot 2^{q_2})/2$, since there are two atoms in each equivalence class where either $k \neq c$ or ($k = c$ and $e \neq d$), and only one atom in the $2^q \cdot 2^{q_2}$ many $\sim$-equivalence classes that contain atoms where $k = c$ and $e = d$.

It will be convenient to introduce notation for these equivalence classes; we shall denote them by $\Gamma_1, \dots, \Gamma_p$, and assume that they are ordered by the least number atom they contain (so that Γ_1 contains γ_1). From above, it follows that each class is

$$\left\{ \begin{pmatrix} k & e \\ d & c \end{pmatrix}, \begin{pmatrix} c & d \\ e & k \end{pmatrix} \right\} \tag{5.10}$$

for some k, c, e, d, and it has either two elements, or just one (when $k = c$ and $e = d$). For fixed k and c, $A(k, c)$ will denote the set of all j such that Γ_j consists of the atoms (5.10) for some e, d.

Note that for $\Theta(b_1, \dots, b_m) = \bigwedge_{1 \leq i,t \leq m} \gamma_{h_{i,t}}(b_i, b_t)$ to be consistent, $[k, c, e, d]$ and $[c, k, d, e]$ must appear exactly the same number of times among the $h_{i,t}$, because $\Theta \models \gamma_{[k,c,e,d]}(a_i, a_t) \iff \Theta \models \gamma_{[c,k,d,e]}(a_t, a_i)$ for these i, t. So when Θ is as in (5.7), since we are considering just the pairs $\langle i, t \rangle \in \{1, \dots, m\}^2$ where $i < t$, only one of $[k, c, e, d]$ and $[c, k, d, e]$ will appear amongst the $h_{i,t}$ for these i, t.

Within the equivalence class (5.10), the unary trace of an atom determines the atom, except when $k = c$ and $e \neq d$, since then $\gamma_{[k,k,e,d]}$ and $\gamma_{[k,k,d,e]}$ are different atoms with the same unary trace. We shall associate a number with each class Γ_j accordingly: 1 if the unary traces do determine its atoms and 2 otherwise. We denote this number s_j.

Definition 5.3. The *signature* of a state description

$$\Theta(b_1,\dots,b_m) = \bigwedge_{1\le i\le m} \beta_{k_i}(b_i) \wedge \bigwedge_{1\le i<t\le m} \gamma_{h_{i,t}}(b_i,b_t)$$

is defined to be the vector $\langle n_1,\dots,n_p\rangle$, where n_j is the number of $\langle i,t\rangle$ such that $1\le i<t\le m$ and $\gamma_{h_{i,t}}\in\Gamma_j$. If Θ is represented by (5.5) and Γ_j is (5.10), then n_j is the number of times one of the atoms from (5.10) appears as a submatrix of (5.5).

We shall define also the *extended signature* of Θ to be

$$\vec{m}\vec{n} = \langle m_1,\dots,m_{2^q}\,;\, n_1,\dots,n_p\rangle,$$

where m_k is the number of times that k appears amongst the k_i, $i=1,\dots,m$.

We remark that the extended signature is derivable from the signature, but it will be convenient for us to record the $\vec{m}$ part explicitly.

Note that if $\vec{m}\vec{n}$ is the extended signature of some state description $\Theta(b_1,\dots,b_m)$ then

$$\sum_{k=1}^{2^q} m_k = m\,, \tag{5.11}$$

for $k\neq c$

$$\sum_{j\in A(k,c)} n_j = m_k m_c\,, \tag{5.12}$$

and

$$\sum_{j\in A(k,k)} n_j = \frac{m_k(m_k-1)}{2}\,. \tag{5.13}$$

The first equation is clear. The reason for the second equation is that for $k\neq c$, if β_k and β_c appear m_k and m_c many times in $\Theta(b_1,\dots,b_m)$ respectively, then m_km_c many of the $h_{i,t}$ in (5.7) will be from an equivalence class in $A(k,c)$ when $i<t$. Otherwise we would have to count $h_{i,t}=[k,c,e,d]$ *and* $h_{t,i}=[c,k,d,e]$ and the sum in (5.12) would be $2m_km_c$. A similar argument works for the third equation.

Conversely, thinking about state descriptions in terms of matrices as in (5.5), we can see that any $\vec{m}\vec{n}=\langle m_1,\dots,m_{2^q}\,;\,n_1,\dots,n_p\rangle$ such that (5.12) and (5.13) hold, is an extended signature of some $\Theta(b_1,\dots,b_m)$ for m defined by (5.11), so we refer to such vectors as *extended signatures on m*.

If the binary case behaved like the unary, Ex would be equivalent to the requirement that the probability of a state description depends only on its signature. However, as we shall see below, this is not the case and so we are led to define the

Signature Exchangeability Principle (binary), BEx

Let L be a binary language and let w be a probability function on SL. Then the probability of a state description depends only on its signature.

BEx implies Ex, since a state description's signature is invariant under permutations of constants. To see this, let $\Theta(b_1, \dots, b_m)$ be a state description as in (5.7), and suppose it has the extended signature $\vec{m}\vec{n}$. Consider

$$\Theta(b_{\tau(1)}, \dots, b_{\tau(m)}) = \bigwedge_{i=1}^{m} \beta_{k_{\tau^{-1}(i)}}(b_i) \wedge \bigwedge_{1 \leq i < t \leq m} \gamma_{h_{\tau^{-1}(i), \tau^{-1}(t)}}(b_i, b_t),$$

the state description obtained from Θ by permuting $b_1, \dots, b_m$ according to $\tau \in \mathsf{S}_m$. Then the extended signature of $\Theta(b_{\tau(1)}, \dots, b_{\tau(m)})$ is also $\vec{m}\vec{n}$. Notice that if for some $1 \leq i < t \leq m$ we have $\Theta(b_{\tau(1)}, \dots, b_{\tau(m)}) \models \gamma_{h_{\tau^{-1}(i), \tau^{-1}(t)}}(b_i, b_t)$ then either $\tau^{-1}(i) < \tau^{-1}(t)$ and $\Theta(b_1, \dots, b_m) \models \gamma_{h_{\tau^{-1}(i), \tau^{-1}(t)}}(b_{\tau^{-1}(i)}, b_{\tau^{-1}(t)})$, or $\tau^{-1}(i) > \tau^{-1}(t)$ and $\Theta(b_1, \dots, b_m) \models \gamma_{h_{\tau^{-1}(t), \tau^{-1}(i)}}(b_{\tau^{-1}(t)}, b_{\tau^{-1}(i)})$. But if $\gamma_{h_{\tau^{-1}(i), \tau^{-1}(t)}} \in \Gamma_j$, then so is $\gamma_{h_{\tau^{-1}(t), \tau^{-1}(i)}}$, so in either case, the same number of atoms from each equivalence class appear in $\Theta(b_1, \dots, b_m)$, $\Theta(b_{\tau(1)}, \dots, b_{\tau(m)})$. Clearly, $\vec{m}$ is the same for both state descriptions since it is derivable from $\vec{n}$ (and in any case, $\{1, \dots, m\} = \{\tau(1), \dots, \tau(m)\}$ as sets).

On the other hand, the converse implication does not hold - two state descriptions with the same signature may not be obtainable from each other by permuting constants and can therefore get different probabilities from functions satisfying Ex. This means that BEx is strictly stronger than Ex. Rather than providing a general proof, we will illustrate why this is so on the case of the language L containing just one binary relation symbol Q. Let $\Theta(b_1, b_2, b_3, b_4)$, $\Phi(b_1, b_2, b_3, b_4)$ be the state descriptions of L represented respectively by the matrices

$$\begin{matrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \qquad\qquad \begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{matrix}\,.$$

Then Θ and Φ have the same signature, namely

$$\vec{n} = \langle n_1 = 3, n_2 = 0, n_3 = 0, n_4 = 3, n_5 = 0, n_6 = 0, n_7 = 0, n_8 = 0, n_9 = 0, n_{10} = 0 \rangle$$

where

$$\Gamma_1 = \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}, \qquad \Gamma_4 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

(and $\vec{m} = \langle m_1 = 4, m_2 = 0 \rangle$). However, there is no permutation of b_1, b_2, b_3, b_4 that maps $\Theta(b_1, b_2, b_3, b_4)$ to $\Phi(b_1, b_2, b_3, b_4)$. To see this let $\sigma \in \mathsf{S}_4$. Any matrix representing $\Theta(b_{\sigma(1)}, b_{\sigma(2)}, b_{\sigma(3)}, b_{\sigma(4)})$ must contain a column consisting of three 0s and one 1, that is the column corresponding to $b_{\sigma^{-1}(4)}$. Since Φ contains no such column, there is no permutation σ such that $\Theta(b_{\sigma(1)}, b_{\sigma(2)}, b_{\sigma(3)}, b_{\sigma(4)})$ is equivalent to $\Phi(b_1, b_2, b_3, b_4)$. Furthermore, the probability function $u^{\bar{p},L}$ with $\bar{p} = \langle 0, \frac{1}{2}, \frac{1}{2}, 0, 0, \dots \rangle$ described in Chapter 3 gives these state descriptions different probabilities: $u^{\bar{p},L}(\Theta) > 0$ while $u^{\bar{p},L}(\Phi) = 0$.

The probability function $u^{\bar{p},L}_{\bar{E}}$ together with the state descriptions Θ and Φ from page 62 provide another example of state descriptions with the same signature and a probability function satisfying Ex that assigns them different probabilities. Recall that these Θ and Φ were constructed to have the same p-spectrum. We remark that the p-spectrum and the signature of a state description are indeed connected. Let $\Theta(b_1, \dots, b_m)$ be a state description as in (5.4). Define the signature+ to record how many $\langle i, t \rangle \in \{1, \dots, m\}^2$ are such that $\gamma_{h_{i,t}} \in \Gamma_j$ for each $j \in \{1, \dots, p\}$ and note that the signature+ is derivable from the extended signature. In addition, suppose we split p-spectrums for state descriptions on m constants into classes such that p-spectrums that agree on $\sum_{\gamma_h \in \Gamma_j} n_h$ for each j (that is, they have the same total number of atoms from each Γ_j) are in the same class. Then state descriptions with the same signature+ all have a p-spectrum from the same class.

The probability functions satisfying BEx share a number of properties with those satisfying Ex in the unary case. In particular, there is a large class of relatively simply defined probability functions similar to the unary $w_{\vec{x}}$ (as described on page 65) which satisfy BEx. These functions are characterised by an independence principle similar to the Constant Independence Principle IP. In addition, we will show that there is a de Finetti-style representation theorem telling us that any probability function satisfying BEx can be expressed as a convex combination of these special functions (as an integral). This, in turn, will yield a proof of a binary generalisation of the Principle of Instantial Relevance. In addition, these results were used in [42] to prove a characterisation of a binary Carnap Continuum as the unique functions satisfying a binary generalisation of Johnson's Sufficientness Postulate. We begin with independence.

5.4 Binary Independence

The Constant Independence Principle IP (for any language), see page 65, requires that any two quantifier free sentences which have no constant symbols in common are stochastically independent. In other words, probability functions satisfying this principle have the property that evidence concerning certain individuals has no impact on probabilities assigned to sentences involving different individuals.

In sentences involving only unary predicate symbols, occurrences of predicates are instantiated by single constants; no predicate can bring two constants together in the way binary relations do. Hence, when the language is unary, the notion of independence used in IP is the strongest one, based on requiring that individuals do not interfere with others. In the binary case, however, beyond simply requiring that individuals do not interfere, we may require the same of pairs of individuals in the following sense.

Definition 5.4. For a sentence ψ of a binary language L we define C^2_ψ to be the set of (unordered) pairs of constants $\{a_i, a_j\}$, $i \neq j$, such that for some binary relation symbol Q of L, either $\pm Q(a_i, a_j)$ or $\pm Q(a_j, a_i)$ appears in ψ.

We say that sentences ϕ, ψ such that C^2_ϕ and C^2_ψ are disjoint *instantiate no pairs in common.* Such sentences cannot reasonably be required to be independent outright because of information each may contain concerning single individuals, but they can be independent conditionally.

Strong Independence Principle (binary), BIP

Let L be a binary language and assume that $\phi, \psi \in QFSL$ instantiate no pairs in common. Let $b_1, \ldots, b_s$ be the constants that ϕ and ψ have in common (if any) and let $\Delta(b_1, \ldots, b_s)$ be a unary trace for these constants. Then

$$w(\phi \wedge \psi \mid \Delta) \;=\; w(\phi \mid \Delta) \cdot w(\psi \mid \Delta). \tag{5.14}$$

If $s = 0$ (the sentences have no constant symbols in common) then $\Delta = \top$ (tautology), so *BIP implies IP*. If Δ is not consistent with ϕ or ψ, or if[4] $w(\Delta) = 0$, then (5.14) clearly holds because both sides are 0.

The Functions $w_{\vec{Y}}$

We shall now define the binary versions $w_{\vec{Y}}$ of the unary $w_{\vec{x}}$ mentioned on page 65. Let $\mathbb{D}_L$ be the set of all

$$\vec{Y} = \langle x_1, \ldots, x_{2^q}\,;\, y_1, \ldots, y_p\rangle$$

such that $x_k, y_j \geq 0$ and $\sum_{k=1}^{2^q} x_k = 1$, and such that for any $1 \leq k, c \leq 2^q$,

$$\sum_{j\,\in A(k,c)} s_j y_j = 1 \tag{5.15}$$

($A(k,c)$ was defined on page 72).[5] We intend to define $w_{\vec{Y}}$ so that these functions satisfy Ex, BIP, $w_{\vec{Y}}(\beta_k(a_i)) = x_k$ and if γ_h is the atom $\gamma_{[k,c,e,d]}$ and Γ_j its equivalence class - that is, Γ_j is (5.10) - then

$$w_{\vec{Y}}(\gamma_h(a_i, a_t) \mid \beta_k(a_i) \wedge \beta_c(a_t)) = y_j\,.$$

To this end, it is convenient to write $j(h)$ for j such that $\gamma_h \in \Gamma_j$. To make the notation more manageable, we also write z_h for $y_{j(h)}$. Hence the y_j are associated with the equivalence classes Γ_j of atoms, and the z_h assign these same values to the individual atoms in these classes. In terms of the z_h, (5.15) says that the sum over z_h for those γ_h with a given unary trace is 1. That is, identifying h with $[k,c,e,d]$, (5.15)

[4] using our convention for conditional probabilities from page 13.

[5] Note that $\mathbb{D}_L$ is both compact and convex in $\mathbb{R}^M$ for a sufficiently large $M \in \mathbb{N}$, a fact we will use in the proof of the representation theorem in Section 5.5.

is equivalent to requiring that

$$\sum_{e,d\in\{1,\ldots,2^{q_2}\}} z_{[k,c,e,d]} = 1. \tag{5.16}$$

To aid with explanations, we will use both (5.15) and (5.16) in what follows.

For a state description

$$\Theta(a_1,\ldots,a_m) = \bigwedge_{1\le i\le m} \beta_{k_i}(a_i) \wedge \bigwedge_{1\le i<t\le m} \gamma_{h_{i,t}}(a_i,a_t) \tag{5.17}$$

we define

$$w_{\vec{Y}}(\Theta(a_1,\ldots,a_m)) = \prod_{1\le i\le m} x_{k_i} \prod_{1\le i<t\le m} z_{h_{i,t}}, \tag{5.18}$$

where $k_i \in \{1,\ldots,2^q\}$ and $z_{h_{i,t}}$ is y_j for that $j \in \{1,\ldots,p\}$ such that $\gamma_{h_{i,t}} \in \Gamma_j$. We take the empty product to be 1. So for instance, for Θ a state formula on one variable, $w_{\vec{Y}}(\Theta(a_1)) = x_{k_1} \cdot 1 = x_{k_1}$.

Proposition 5.5. *Let L be a binary language. The functions $w_{\vec{Y}}$ defined in* (5.18) *determine probability functions on SL that satisfy Ex.*

Proof. $w_{\vec{Y}}$ is defined on state descriptions so it suffices to check that conditions (P1′)-(P3′) from page 13 hold. (P1′) clearly holds. To check that (P2′) holds, consider the state description $\Theta_i(a_1)$ and notice that $w_{\vec{Y}}(\Theta_i(a_1)) = x_i$ for some $i = 1,\ldots,2^q$. Then since $\models \bigvee_{i\in\{1,\ldots,2^q\}} \Theta_i(a_1)$, we have that

$$w_{\vec{Y}}(\top) = w_{\vec{Y}}\left(\bigvee_{i\in\{1,\ldots,2^q\}} \Theta_i(a_1)\right) = \sum_{i=1}^{2^q} w_{\vec{Y}}(\Theta_i(a_1)) = \sum_{i=1}^{2^q} x_i = 1$$

since the state formulae $\Theta_1(x),\ldots,\Theta_{2^q}(x)$ are mutually exclusive. Alternatively, recall our convention from page 12 that the empty state description is a tautology. Then since we defined the empty product to be 1, it follows that for any tautology $\top$, $w_{\vec{Y}}(\top) = 1$ so (P2′) is satisfied.

To see that (P3′) holds, note that whenever $\Phi(a_1,\ldots,a_m,a_{m+1}) \models \Theta(a_1,\ldots,a_m)$ we have

$$\Phi(a_1,\ldots,a_m,a_{m+1}) \equiv \Theta(a_1,\ldots,a_m) \wedge \beta_{k_{m+1}}(a_{m+1}) \wedge \bigwedge_{i\in\{1,\ldots,m\}} \gamma_{h_{i,m+1}}(a_i,a_{m+1})$$

where the unary trace of $\gamma_{h_{i,m+1}}(x,y)$ is $\beta_{k_i}(x) \wedge \beta_{k_{m+1}}(y)$, and hence using (5.18)

$$\begin{aligned} w_{\vec{Y}}(\Phi) &= w_{\vec{Y}}\left(\Theta \wedge \beta_{k_{m+1}}(a_{m+1}) \wedge \bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\right) \\ &= w_{\vec{Y}}(\Theta) \cdot x_{k_{m+1}} \cdot \prod_{i\in\{1,\dots,m\}} z_{h_{i,m+1}}. \end{aligned}$$

Consider $\beta_{k_{m+1}}(a_{m+1}) \wedge \bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})$. We may choose any $k_{m+1} \in \{1,\dots,2^q\}$ and Φ will remain consistent with Θ. Furthermore, for each such choice of k_{m+1} and each $i \in \{1,\dots,m\}$, we may choose any of the $\gamma_{h_{i,m+1}}$ that have unary trace $\beta_{k_i}(x) \wedge \beta_{k_{m+1}}(y)$.

Therefore for Θ, Φ as above

$$\begin{aligned} \sum_{\Phi \models \Theta} w_{\vec{Y}}(\Phi) &= \sum_{\Phi \models \Theta} \left(w_{\vec{Y}}(\Theta) \cdot x_{k_{m+1}} \cdot \prod_{i\in\{1,\dots,m\}} z_{h_{i,m+1}} \right) \\ &= w_{\vec{Y}}(\Theta) \cdot \sum_{\Phi \models \Theta} \left(x_{k_{m+1}} \cdot \prod_{i\in\{1,\dots,m\}} z_{h_{i,m+1}} \right) \\ &= w_{\vec{Y}}(\Theta) \cdot \sum_{c=1}^{2^q} \left(x_c \cdot \sum_{\vec{e},\vec{d}\in\{1,\dots,2^{q_2}\}^m} \prod_{i\in\{1,\dots,m\}} z_{[k_i,c,e_i,d_i]} \right) \\ &= w_{\vec{Y}}(\Theta) \cdot \sum_{c=1}^{2^q} \left(x_c \cdot \prod_{i\in\{1,\dots,m\}} \sum_{e,d\in\{1,\dots,2^{q_2}\}} z_{[k_i,c,e,d]} \right) \end{aligned}$$

where $\vec{e} = \langle e_1, e_2, \dots, e_m \rangle$, $\vec{d} = \langle d_1, d_2, \dots, d_m \rangle$ and e_i, d_i are the ith entry of $\vec{e}, \vec{d}$ respectively. By the definition of $\vec{Y}$, $\sum_{e,d\in\{1,\dots,2^{q_2}\}} z_{[k_i,c,e,d]} = 1$ and $\sum_{c=1}^{2^q} x_c = 1$, so

$$\sum_{\Phi \models \Theta} w_{\vec{Y}}(\Phi) = w_{\vec{Y}}(\Theta) \cdot \sum_{c=1}^{2^q} x_c \cdot 1 = w_{\vec{Y}}(\Theta).$$

Hence (P3′) holds, too. It follows that $w_{\vec{Y}}$ extends to a probability function on SL.

We now show that $w_{\vec{Y}}$ satisfies Ex. If $\Theta(a_1,\dots,a_m)$ is as in (5.17), $\sigma \in \mathsf{S}_m$ and

$$\Psi(a_1,\dots,a_m) = \Theta(a_{\sigma(1)},\dots,a_{\sigma(m)}),$$

then[6]

$$\Psi(a_1,\dots,a_m) = \bigwedge_{1\le i\le m} \beta_{k_{\sigma^{-1}(i)}}(a_i) \wedge \bigwedge_{1\le i<t\le m} \gamma_{h_{\sigma^{-1}(i),\sigma^{-1}(t)}}(a_i, a_t),$$

[6]Recall the convention from page 69 needed below when $\sigma^{-1}(i) > \sigma^{-1}(t)$.

and using a similar argument to that on page 74, the multiset $\{k_{\sigma^{-1}(i)} : 1 \leq i \leq m\}$ equals the multiset $\{k_i : 1 \leq i \leq m\}$, and the multisets $\{h_{\sigma^{-1}(i),\sigma^{-1}(t)} : 1 \leq i < t \leq m\}$, $\{h_{i,t} : 1 \leq i < t \leq m\}$ can only differ in that the former contains $h_{i',t'}$ in place of $h_{t',i'}$ when $i' = \sigma^{-1}(i) > \sigma^{-1}(t) = t'$. We have $\gamma_{h_{i',t'}} \sim \gamma_{h_{t',i'}}$, i.e. they are in the same equivalence class Γ_j, hence $z_{h_{i',t'}} = z_{h_{t',i'}}$ (since both are y_j for this j). Therefore

$$w_{\vec{Y}}(\Psi) = \prod_{1 \leq i \leq m} x_{k_{\sigma^{-1}(i)}} \prod_{1 \leq i < t \leq m} z_{h_{\sigma^{-1}(i),\sigma^{-1}(t)}} = \prod_{1 \leq i \leq m} x_{k_i} \prod_{1 \leq i < t \leq m} z_{h_{i,t}} = w_{\vec{Y}}(\Theta).$$

Ex now follows, since (using for example, [36, Lemma 7.3]), if for any $m \in \mathbb{N}^+$, for any permutation $\sigma \in \mathsf{S}_m$ and state description $\Theta(a_1, \ldots, a_m)$, $w(\Theta(a_1, \ldots, a_m)) = w(\Theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)}))$, then w satisfies Ex. □

Proposition 5.6. *The $w_{\vec{Y}}$ satisfy BEx.*

Proof. Notice that using a similar argument to above, (5.18) remains valid even when we replace the $a_1, \ldots, a_m$ by other distinct constants $b_1, \ldots, b_m$. We sum the probabilities of state descriptions for $a_1, \ldots, a_M$ extending $\Theta(b_1, \ldots, b_m)$, where M is sufficiently large so that all the $b_1, \ldots, b_m$ are amongst the $a_1, \ldots, a_M$. For the a_i that do not appear amongst $b_1, \ldots, b_m$ there is a free choice which atoms they satisfy, so summing over all possible $\Phi(a_1, \ldots, a_M) \models \Theta(b_1, \ldots, b_m)$ means summing over all possible x_{k_i} factors and all possible z_h for γ_h with some fixed unary trace, both of which sum up to 1, thus not affecting $w_{\vec{Y}}(\Theta)$. Therefore $w_{\vec{Y}}(\Theta)$ is independent of the choice of constants instantiating Θ, and hence $w_{\vec{Y}}$ satisfies BEx since the right hand side of (5.18) depends only on the signature of Θ. □

We now show how the binary versions of the familiar probability functions c_0^L and c_∞^L from Carnap's Continuum[7] can be produced using the $w_{\vec{Y}}$ for a binary language L. These functions feature extensively in studies of Inductive Logic, so the fact they can be formed from the $w_{\vec{Y}}$ supports in some sense the credibility of the $w_{\vec{Y}}$. Consider $\vec{Y}$ defined by

$$x_1 = x_2 = \cdots = x_{2^q} = 2^{-q},$$

[7]See page 17.

$$y_1 = y_2 = \cdots = y_p = 2^{-2q_2}.$$

Then $\vec{Y} \in \mathbb{D}_L$ since $\sum_{1 \le k \le 2^q} x_k = 1$ and, identifying again $z_h = z_{[k,c,e,d]}$ with $y_{j(h)}$ as defined on page 77, $\sum_{e,d \in \{1,\ldots,2^{q_2}\}} z_{[k,c,e,d]} = \sum_{e,d \in \{1,\ldots,2^{q_2}\}} 2^{-2q_2} = 1$. Furthermore, $w_{\vec{Y}}$ defined with this $\vec{Y}$ treats each $\pm P_s(a_i)$ and each $\pm Q_u(a_i, a_t)$ (where a_i, a_t are not necessarily distinct) as stochastically independent and each occurring with probability $\frac{1}{2}$, which gives c_∞^L. To see that each instantiated relation symbol occurs with probability $\frac{1}{2}$, notice that for this $\vec{Y}$, the probability of $\pm P_s(a_i)$ is the sum of the probabilities for exactly half of the β_k, and the probability of $\pm Q_u(a_i, a_t)$ is the sum of the probabilities of half of the γ_h. That is, assuming the trace of $\gamma_h(x, y)$ is $\beta_k(x) \wedge \beta_c(y)$, the probability of each $\gamma_h(a_i, a_t)$ is

$$x_k \cdot x_c \cdot z_h = 2^{-q} \cdot 2^{-q} \cdot 2^{-2q_2},$$

so the probability of $\pm Q_u(a_i, a_t)$ is

$$\frac{N}{2}(x_k \cdot x_c \cdot z_h) = 2^{2q+2q_2-1}(2^{-q} \cdot 2^{-q} \cdot 2^{-2q_2}) = \frac{1}{2}$$

where $N = 2^{2q} 2^{2q_2}$ is the total number of atoms.

To see that this $w_{\vec{Y}}$ treats instantiations of relation symbols as stochastically independent, let ψ be the sentence

$$\pm P_{s_1}(a_{i_1}) \wedge \pm P_{s_2}(a_{i_2}) \wedge \ldots \wedge \pm P_{s_{n_1}}(a_{i_{n_1}}) \\ \wedge \pm Q_{u_1}(a_{i_1}, a_{t_1}) \wedge \pm Q_{u_2}(a_{i_2}, a_{t_2}) \wedge \ldots \wedge \pm Q_{u_{n_2}}(a_{i_{n_2}}, a_{t_{n_2}})$$

for some assignment of $\pm$ and where $n_1 + n_2 = n$, $s_1, \ldots, s_{n_1}$ are from $\{1, \ldots, q_1\}$, and $u_1, \ldots, u_{n_2}$ are from $\{1, \ldots, q_2\}$. Let $a_1, \ldots, a_m$ be such that all of $a_{i_1}, \ldots, a_{i_{n_1}}, a_{t_1}, \ldots, a_{t_{n_2}}$ are included amongst them. Then

$$w_{\vec{Y}}(\psi) = \sum_{\Theta(a_1,\ldots,a_m) \models \psi} w_{\vec{Y}}(\Theta) = \sum_{\Theta(a_1,\ldots,a_m) \models \psi} w_{\vec{Y}}\left(\bigwedge_{1 \le i \le m} \beta_{k_i}(a_i) \wedge \bigwedge_{1 \le i < t \le m} \gamma_{h_{i,t}}(a_i, a_t) \right) \\ = \sum_{\Theta(a_1,\ldots,a_m) \models \psi} \left(\prod_{1 \le i \le m} x_{k_i} \prod_{1 \le i < t \le m} z_{h_{i,t}} \right) = \sum_{\Theta(a_1,\ldots,a_m) \models \psi} (2^{-q})^m \cdot (2^{-2q_2})^{\frac{1}{2}m(m-1)}$$

since each x_{k_i} is 2^{-q} and there are m of these, and since there are $\frac{1}{2}m(m-1)$ pairs $\langle a_i, a_t \rangle$ such that $1 \le i < t \le m$ and $z_{h_{i,t}}$ for each of them is 2^{-2q_2}. In addition, there are $M = (2^{q_1})^m \cdot (2^{q_2})^{m^2}$ possible state descriptions for $a_1, \ldots, a_m$. Of these, half logically imply the assignment of $\pm P_{s_1}(a_{i_1})$, of those half satisfy $\pm P_{s_2}(a_{i_2})$ and so on, so there are $M \cdot \left(\frac{1}{2}\right)^n$ state descriptions $\Theta(a_1, \ldots, a_m)$ such that $\Theta \models \psi$. It follows that the probability $w_{\vec{Y}}(\psi)$ is

$$\begin{aligned}\Big((2^{q_1})^m \cdot (2^{q_2})^{m^2} \cdot 2^{-n}\Big) \cdot \left(2^{-q}\right)^m \cdot \left(2^{-2q_2}\right)^{\frac{1}{2}m(m-1)} \\ = \Big((2^{q_1})^m \cdot (2^{q_2})^{m^2} \cdot 2^{-n}\Big) \cdot (2^{-q_1}2^{-q_2})^m \cdot (2^{-2q_2})^{\frac{1}{2}m(m-1)} = 2^{-n},\end{aligned}$$

as required.

To produce c_0^L, recall that this function gives non-zero probability only to state descriptions in which all the constants behave in the same way. Let

$$\beta_k(x) = \bigwedge_{s=1}^{q_1} \pm P_s(x) \wedge \bigwedge_{u=1}^{q_2} \pm Q_u(x,x)$$

for some assignment of $\pm$, and let $e \in \{1, \ldots, 2^{q_2}\}$ be such that

$$\beta_k(x) \models Q_u(x,x) \iff \delta_e(x,y) \models Q_u(x,y).$$

Denote this e by $e(k)$.

Let $\gamma_h \in \Gamma_j$ be the atom $\gamma_{[k,k,e(k),e(k)]}$. Define $\vec{Y_k} = \langle x_1, \ldots, x_{2^q}\,;\, y_1, \ldots, y_p\rangle$ to be such that $x_k = 1$, $y_j = 1$ for this j (the rest of the x_s are 0 and all other[8] y_j are such that $\vec{Y}$ satisfies (5.15)). Then $\vec{Y_k} \in \mathbb{D}_L$ and we can define $w_{\vec{Y_k}}(\Theta)$ for a state description Θ. There are 2^q atoms of the form $\gamma_{[k,k,e(k),e(k)]}$ since there are 2^q choices of k and each one determines such an atom. Clearly, the individuals satisfying this atom behave in the same way, so if all constants of a state description satisfy this atom, this state description should be given a non-zero probability by c_0^L. Furthermore, if γ_h is not $\gamma_{[k,k,e(k),e(k)]}$ for some k, then for any state description Θ, if $\Theta \models \gamma_h(b_i, b_t)$ for some individuals b_i, b_t, then $c_0^L(\Theta) = 0$. Therefore, the value given by c_0^L to a state description can be given as the convex sum of the $w_{\vec{Y_k}}$ - that is -

$$c_0^L(\Theta) = 2^{-q} \sum_{k=1}^{2^q} w_{\vec{Y_k}}(\Theta).$$

We would now like to show that the $w_{\vec{Y}}$ satisfy BIP. For this purpose, we first need the following lemma regarding $w_{\vec{Y}}$ acting on partial state descriptions of L.

[8]Note that the values of the other y_j have no effect, since if the trace of $\gamma_h(x,y)$ is $\beta_k(x) \wedge \beta_c(y)$ and one of x_k, x_c is 0, then this $y_{j(h)}$ will not contribute to $w_{\vec{Y}}(\Theta)$ for any Θ.

Lemma 5.7. *Let L be a binary language and let*

$$\Phi(b_1,\dots,b_m)=\bigwedge_{1\le i\le m}\beta_{k_i}(b_i)\wedge\bigwedge_{\substack{\{b_i,b_t\}\in C\\ i<t}}\gamma_{h_{i,t}}(b_i,b_t) \tag{5.19}$$

be a partial state description of L. Then

$$w_{\vec{Y}}(\Phi(b_1,\dots,b_m))=\prod_{1\le i\le m}x_{k_i}\prod_{\substack{\{b_i,b_t\}\in C\\ i<t}}z_{h_{i,t}}.$$

Proof. Let $\Theta(b_1,\dots,b_m)$ be a (complete) state description such that $\Theta\models\Phi$. Then

$$\begin{aligned}
\Theta(b_1,\dots,b_m) &= \bigwedge_{1\le i\le m}\beta_{k_i}(b_i)\wedge\bigwedge_{1\le i<t\le m}\gamma_{h_{i,t}}(b_i,b_t)\\
&= \bigwedge_{1\le i\le m}\beta_{k_i}(b_i)\wedge\bigwedge_{\substack{\{b_i,b_t\}\in C\\ i<t}}\gamma_{h_{i,t}}(b_i,b_t)\wedge\bigwedge_{\substack{1\le i<t\le m\\ \{b_i,b_t\}\notin C}}\gamma_{h_{i,t}}(b_i,b_t)\\
&= \Phi(b_1,\dots,b_m)\wedge\bigwedge_{\substack{1\le i<t\le m\\ \{b_i,b_t\}\notin C}}\gamma_{h_{i,t}}(b_i,b_t).
\end{aligned}$$

So we have

$$\begin{aligned}
w_{\vec{Y}}(\Phi(b_1,\dots,b_m)) &= \sum_{\Theta\models\Phi}w_{\vec{Y}}(\Theta(b_1,\dots,b_m))\\
&= \sum_{\Theta\models\Phi}\left(\prod_{1\le i\le m}x_{k_i}\cdot\prod_{\substack{\{b_i,b_t\}\in C\\ i<t}}z_{h_{i,t}}\cdot\prod_{\substack{1\le i<t\le m\\ \{b_i,b_t\}\notin C}}z_{h_{i,t}}\right)\\
&= \prod_{1\le i\le m}x_{k_i}\cdot\prod_{\substack{\{b_i,b_t\}\in C\\ i<t}}z_{h_{i,t}}\cdot\sum_{\Theta\models\Phi}\left(\prod_{\substack{1\le i<t\le m\\ \{b_i,b_t\}\notin C}}z_{h_{i,t}}\right)\\
&= \prod_{1\le i\le m}x_{k_i}\cdot\prod_{\substack{\{b_i,b_t\}\in C\\ i<t}}z_{h_{i,t}}\cdot\sum_{\substack{e_{i,t},\ d_{i,t}\in\{1,\dots,2^{q_2}\}\\ 1\le i<t\le m\\ \{b_i,b_t\}\notin C}}\left(\prod_{\substack{1\le i<t\le m\\ \{b_i,b_t\}\notin C}}z_{[k_i,k_t,e_{i,t},d_{i,t}]}\right)\\
&= \prod_{1\le i\le m}x_{k_i}\cdot\prod_{\substack{\{b_i,b_t\}\in C\\ i<t}}z_{h_{i,t}}\cdot\prod_{\substack{1\le i<t\le m\\ \{b_i,b_t\}\notin C}}\left(\sum_{e,d\in\{1,\dots,2^{q_2}\}}z_{[k_i,k_t,e,d]}\right)\\
&= \prod_{1\le i\le m}x_{k_i}\cdot\prod_{\substack{\{b_i,b_t\}\in C\\ i<t}}z_{h_{i,t}},
\end{aligned}$$

where the third equality follows since every Θ that extends Φ would have these two products in common; the fourth equality is because for each pair $\{b_i,b_t\}\notin C$ with

$1 \leq i < t \leq m$, we can choose for $\gamma_{h_{i,t}}$ any atom with unary trace $\beta_{k_i}(x) \wedge \beta_{k_t}(y)$ and we will get a state description that extends Φ, and the final equality is since $\sum_{e,d \in \{1,\dots,2^{q_2}\}} z_{[k_i,k_t,e,d]} = 1$ for any fixed k_i, k_t from $\{1, \dots, 2^q\}$. $\square$

Theorem 5.8. *Let L be a binary language. The probability functions $w_{\vec{Y}}$ on SL satisfy BIP and hence IP.*

Furthermore, any probability function on SL satisfying Ex and BIP is equal to $w_{\vec{Y}}$ for some $\vec{Y} \in \mathbb{D}_L$.

Proof. Assume that Φ and Ψ are some partial state descriptions which instantiate no pairs in common. Let $b_1, \dots, b_s$ be the constants that Φ and Ψ have in common and let Δ be a unary trace for these constants. If Δ is not consistent with Φ or Ψ, then

$$w_{\vec{Y}}(\Phi \wedge \Psi \mid \Delta) = w_{\vec{Y}}(\Phi \mid \Delta)\, w_{\vec{Y}}(\Psi \mid \Delta) \tag{5.20}$$

holds because both sides are 0, as remarked on page 77. So suppose Φ is as in (5.19), $s \leq m$,

$$\Psi(b_1, \dots, b_s, b_{m+1}, \dots, b_{m+n}) = \bigwedge_{1 \leq i \leq s} \beta_{k_i}(b_i) \wedge \bigwedge_{m+1 \leq i \leq m+n} \beta_{k_i}(b_i) \wedge \bigwedge_{\substack{\{b_i,b_t\} \in D \\ i<t}} \gamma_{h_{i,t}}(b_i, b_t) \tag{5.21}$$

where D is some set of 2-element subsets of $\{b_1, \dots, b_s, b_{m+1}, \dots, b_{m+n}\}$, $D \cap C = \emptyset$, and

$$\Delta(b_1, \dots, b_s) = \bigwedge_{1 \leq i \leq s} \beta_{k_i}(b_i).$$

We can now use the above lemma regarding values of $w_{\vec{Y}}$ for partial state descriptions to prove that (5.20) holds in this case, too: $\Theta \wedge \Phi$ is again a partial state description, so the left hand side of (5.20) is

$$\begin{aligned} w_{\vec{Y}}(\Phi \wedge \Psi \mid \Delta) = \frac{w_{\vec{Y}}(\Phi \wedge \Psi)}{w_{\vec{Y}}(\Delta)} &= \frac{\prod_{1 \leq i \leq m+n} x_{k_i} \prod_{\substack{\{b_i,b_t\} \in C \cup D \\ i<t}} z_{h_{i,t}}}{\prod_{1 \leq i \leq s} x_{k_i}} \\ &= \prod_{s+1 \leq i \leq m+n} x_{k_i} \prod_{\substack{\{b_i,b_t\} \in C \cup D \\ i<t}} z_{h_{i,t}}. \end{aligned}$$

The right hand side of (5.20) is $w_{\vec{Y}}(\Phi \mid \Delta)\, w_{\vec{Y}}(\Psi \mid \Delta)$ which is

$$\begin{aligned}
\frac{w_{\vec{Y}}(\Phi)}{w_{\vec{Y}}(\Delta)} \cdot \frac{w_{\vec{Y}}(\Psi)}{w_{\vec{Y}}(\Delta)} &= \frac{\prod_{1\le i\le m} x_{k_i} \prod_{\substack{\{b_i,b_t\}\in C\\ i<t}} z_{h_{i,t}}}{\prod_{1\le i\le s} x_{k_i}} \cdot \frac{\prod_{1\le i\le s} x_{k_i} \prod_{m+1\le i\le m+n} x_{k_i} \prod_{\substack{\{b_i,b_t\}\in D\\ i<t}} z_{h_{i,t}}}{\prod_{1\le i\le s} x_{k_i}} \\
&= \prod_{s+1\le i\le m} x_{k_i} \prod_{\substack{\{b_i,b_t\}\in C\\ i<t}} z_{h_{i,t}} \cdot \prod_{m+1\le i\le m+n} x_{k_i} \prod_{\substack{\{b_i,b_t\}\in D\\ i<t}} z_{h_{i,t}} \\
&= \prod_{s+1\le i\le m+n} x_{k_i} \prod_{\substack{\{b_i,b_t\}\in C\cup D\\ i<t}} z_{h_{i,t}}.
\end{aligned}$$

So both sides of (5.20) are equal. Hence BIP holds when ϕ, ψ are partial state descriptions.

To prove that (5.14) holds with general $\phi, \psi \in QFSL$, note that any quantifier free sentence is equivalent to a disjunction of partial state descriptions by a slight adaptation of the usual proof of the Disjunctive Normal Form Theorem (swapping propositions for instantiations of relations). So suppose $\phi(b_1, \ldots, b_m) \in QFSL$ is equivalent to a disjunction of partial state descriptions Φ_u as in (5.19), with $C = C^2_\phi$. Assume that $\psi \in QFSL$ instantiates no pairs in common with ϕ. Without loss of generality, let $b_1, \ldots, b_s$ be the constants that ϕ and ψ have in common and $b_{m+1}, \ldots, b_{m+n}$ the remaining constants appearing in ψ. ψ is equivalent to a disjunction of partial state descriptions Ψ_f as in (5.21) where $D = C^2_\psi$, and so by the above, for any unary trace Δ for $b_1, \ldots, b_s$,

$$\begin{aligned}
w_{\vec{Y}}(\,\phi \wedge \psi \mid \Delta) = w_{\vec{Y}}\left(\bigvee_u \Phi_u \wedge \bigvee_f \Psi_f \,\middle|\, \Delta\right) &= \sum_{u,f} w_{\vec{Y}}(\,\Phi_u \wedge \Psi_f \mid \Delta) \\
&= \sum_{u,f} w_{\vec{Y}}\,(\Phi_u \mid \Delta) \cdot w_{\vec{Y}}\,(\Psi_f \mid \Delta) = \sum_u w_{\vec{Y}}\,(\Phi_u \mid \Delta) \cdot \sum_f w_{\vec{Y}}\,(\Psi_f \mid \Delta) \\
&= w_{\vec{Y}}\left(\bigvee_u \Phi_u \,\middle|\, \Delta\right) \cdot w_{\vec{Y}}\left(\bigvee_f \Psi_f \,\middle|\, \Delta\right) = w_{\vec{Y}}(\phi \mid \Delta) \cdot w_{\vec{Y}}(\psi \mid \Delta),
\end{aligned}$$

as required, where the equalities follow since the Φ_u and similarly the Ψ_f are mutually exclusive, Φ_u and Ψ_f instantiate no pairs in common for each u, f, and since $w_{\vec{Y}}$ satisfies BIP on partial state descriptions.

So far we have shown that the $w_{\vec{Y}}$ satisfy BIP (and hence also IP). For the final part of the theorem, assume that w satisfies Ex and BIP. We define

$$x_k = w(\beta_k(a_i))$$

and

$$y_{j(h)} = z_h = w(\gamma_h(a_i, a_t) \mid \beta_k(a_i) \wedge \beta_c(a_t))$$

where $\beta_k(x) \wedge \beta_c(y)$ is the unary trace of $\gamma_h(x, y)$. Note that by Ex, this definition is correct in that it does not matter which a_i, a_t we take, and when $j = j(h) = j(g)$ (that is, when $\gamma_h \sim \gamma_g$), then $z_h = z_g$, and y_j is given the same value. Using BIP, we will check that with $\vec{Y}$ defined in this way, $w_{\vec{Y}}$ equals w for state descriptions, and hence $w = w_{\vec{Y}}$ for all sentences. We will do this by showing that $\vec{Y} \in \mathbb{D}_L$, and that $w_{\vec{Y}} = w$ on state descriptions of increasing numbers of constants.

Firstly, notice that x_k, y_j must be non-negative since w is a probability function. Since w satisfies BIP we have

$$w\left(\bigwedge_{i\in S} \beta_{k_i}(a_i)\right) = \prod_{i\in S} x_{k_i}$$

(where S is any finite set), so also for $\gamma_h(x, y)$ from Γ_j with unary trace $\beta_k(x) \wedge \beta_c(y)$

$$w(\gamma_h(a_i, a_t)) = w(\gamma_h(a_i, a_t) \mid \beta_k(a_i) \wedge \beta_c(a_t)) \cdot w(\beta_k(a_i) \wedge \beta_c(a_t)) = x_k\, x_c\, y_j.$$

In addition, $w(\top) = w\left(\bigvee_{k=1}^{2^q} \beta_k(a_1)\right) = \sum_{k=1}^{2^q} x_k = 1$, and the values for $y_{j(h)} = z_h$ are such that (5.16) holds, since

$$\begin{aligned}
1 = w(\top) = w&\left(\bigvee_{k,c\in\{1,\dots,2^q\}} \bigvee_{e,d\in\{1,\dots,2^{q_2}\}} \gamma_{[k,c,e,d]}(a_i, a_t)\right)\\
&= \sum_{k,c\in\{1,\dots,2^q\}} \sum_{e,d\in\{1,\dots,2^{q_2}\}} w(\gamma_{[k,c,e,d]}(a_i, a_t))\\
&= \sum_{k,c\in\{1,\dots,2^q\}} \sum_{e,d\in\{1,\dots,2^{q_2}\}} x_k\, x_c\, z_{[k,c,e,d]}\\
&= \sum_{k,c\in\{1,\dots,2^q\}} \left(x_k\, x_c \sum_{e,d\in\{1,\dots,2^{q_2}\}} z_{[k,c,e,d]}\right)
\end{aligned}$$

and hence $\sum_{e,d\in\{1,\dots,2^{q_2}\}} z_{[k,c,e,d]} = 1$ since $\sum_{k,c\in\{1,\dots,2^q\}} x_k\, x_c = 1$.

So we already have that $\vec{Y} \in \mathbb{D}_L$ and that $w = w_{\vec{Y}}$ on state descriptions of one or two constants. Next consider a state description $\Theta_3(a_1, a_2, a_3)$.[9] Recall that we can write it as the conjunction

$$\gamma_{h_{1,2}}(a_1, a_2) \wedge \gamma_{h_{1,3}}(a_1, a_3) \wedge \gamma_{h_{2,3}}(a_2, a_3)$$

[9] We could jump to the inductive step already at this point, but presenting the argument also for state descriptions of 3 individuals helps to see the general case.

(as in (5.6)), and suppose the unary trace of Θ_3 is $\beta_{k_1}(a_1) \wedge \beta_{k_2}(a_2) \wedge \beta_{k_3}(a_3)$. Let $\psi = \beta_{k_2}(a_2) \wedge \beta_{k_3}(a_3)$. Since w satisfies BIP we have that

$$
\begin{aligned}
& w(\Theta_3(a_1,a_2,a_3) \mid \psi) \\
&= w(\gamma_{h_{1,2}}(a_1,a_2) \wedge \gamma_{h_{1,3}}(a_1,a_3) \mid \psi) \cdot w(\gamma_{h_{2,3}}(a_2,a_3) \mid \psi) \\
&= \left(\frac{w(\gamma_{h_{1,2}}(a_1,a_2) \wedge \gamma_{h_{1,3}}(a_1,a_3))}{x_{k_2}\, x_{k_3}}\right) \cdot \left(\frac{w(\gamma_{h_{2,3}}(a_2,a_3))}{x_{k_2}\, x_{k_3}}\right) \\
&= \left(\frac{w(\gamma_{h_{1,2}}(a_1,a_2) \wedge \gamma_{h_{1,3}}(a_1,a_3) \mid \beta_{k_1}(a_1))}{x_{k_2}\, x_{k_3}}\right) \cdot w(\beta_{k_1}(a_1)) \cdot \left(\frac{x_{k_2}\, x_{k_3}\, z_{h_{2,3}}}{x_{k_2}\, x_{k_3}}\right) \\
&= \left(\frac{w(\gamma_{h_{1,2}}(a_1,a_2) \mid \beta_{k_1}(a_1)) \cdot w(\gamma_{h_{1,3}}(a_1,a_3) \mid \beta_{k_1}(a_1))}{x_{k_2}\, x_{k_3}}\right) \cdot w(\beta_{k_1}(a_1)) \cdot \left(\frac{x_{k_2}\, x_{k_3}\, z_{h_{2,3}}}{x_{k_2}\, x_{k_3}}\right) \\
&= \left(\frac{w(\gamma_{h_{1,2}}(a_1,a_2) \mid \beta_{k_1}(a_1)) \cdot w(\gamma_{h_{1,3}}(a_1,a_3) \mid \beta_{k_1}(a_1))}{x_{k_2}\, x_{k_3}}\right) \cdot x_{k_1}\, z_{h_{2,3}} \\
&= \left(\frac{w(\gamma_{h_{1,2}}(a_1,a_2))}{x_{k_1} x_{k_2}}\right) \cdot \left(\frac{w(\gamma_{h_{1,3}}(a_1,a_3))}{x_{k_1} x_{k_3}}\right) \cdot x_{k_1}\, z_{h_{2,3}} \\
&= \frac{x_{k_1}\, x_{k_2}\, z_{h_{1,2}}}{x_{k_1}\, x_{k_2}} \cdot \frac{x_{k_1}\, x_{k_3}\, z_{h_{1,3}}}{x_{k_1}\, x_{k_3}} \cdot x_{k_1}\, z_{h_{2,3}} \\
&= x_{k_1}\, z_{h_{1,2}}\, z_{h_{1,3}}\, z_{h_{2,3}}\,,
\end{aligned}
$$

where we used the fact that $(\gamma_{h_{1,2}}(a_1,a_2) \wedge \gamma_{h_{1,3}}(a_1,a_3)) \models \beta_{k_1}(a_1)$ and hence $w(\gamma_{h_{1,2}}(a_1,a_2) \wedge \gamma_{h_{1,3}}(a_1,a_3)) = w(\gamma_{h_{1,2}}(a_1,a_2) \wedge \gamma_{h_{1,3}}(a_1,a_3) \mid \beta_{k_1}(a_1)) \cdot w(\beta_{k_1}(a_1))$. So

$$\frac{w(\Theta_3(a_1,a_2,a_3))}{w(\psi)} = x_{k_1}\, z_{h_{1,2}}\, z_{h_{1,3}}\, z_{h_{2,3}}$$

and hence

$$w(\Theta_3(a_1,a_2,a_3)) = \prod_{1 \le i \le 3} x_{k_i} \prod_{1 \le i < t \le 3} z_{h_{i,t}}.$$

Now suppose the same holds for state descriptions of m constants for some $m \in \mathbb{N}^+$, that is, for every Θ_m

$$w(\Theta_m(a_1,\dots,a_m)) = \prod_{1 \le i \le m} x_{k_i} \prod_{1 \le i < t \le m} z_{h_{i,t}}. \tag{5.22}$$

Let $\Theta_{m+1}(a_1,\dots,a_{m+1}) \models \Theta_m(a_1,\dots,a_m)$. We can write Θ_{m+1} in the form

$$\Theta_m(a_1,\dots,a_m) \wedge \beta_{k_{m+1}}(a_{m+1}) \wedge \bigwedge_{i \in \{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i,a_{m+1}). \tag{5.23}$$

Consider first $\bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})$, the rightmost term in (5.23). Using BIP and letting $\psi = \beta_{k_{m+1}}(a_{m+1})$ we have that (since $w(\gamma_{h_{i,m+1}}(a_i, a_{m+1})) = x_{k_i}\, x_{k_{m+1}}\, z_{h_{i,m+1}}$)

$$
\begin{aligned}
w&\left(\bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\right)\\
&= w\big(\gamma_{h_{1,m+1}}(a_1, a_{m+1}) \,|\, \psi\big) \cdot w\left(\bigwedge_{i\in\{2,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1}) \,\middle|\, \psi\right) w(\psi)\\
&= \frac{w\big(\gamma_{h_{1,m+1}}(a_1, a_{m+1})\big)}{w(\psi)} \cdot \frac{w\Big(\bigwedge_{i\in\{2,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\Big)}{w(\psi)} \cdot w(\psi)\\
&= x_{k_1}\, z_{h_{1,m+1}} \cdot w\left(\bigwedge_{i\in\{2,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\right)\\
&= x_{k_1}\, z_{h_{1,m+1}} \cdot w\big(\gamma_{h_{2,m+1}}(a_2, a_{m+1}) \,|\, \psi\big) \cdot w\left(\bigwedge_{i\in\{3,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1}) \,\middle|\, \psi\right) w(\psi)\\
&= x_{k_1}\, x_{k_2}\, z_{h_{1,m+1}}\, z_{h_{2,m+1}} \cdot w\left(\bigwedge_{i\in\{3,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\right)\\
&\;\;\vdots\\
&= \left(\prod_{1\le i\le m-1} x_{k_i} \prod_{i\in\{1,\dots,m-1\}} z_{h_{i,m+1}}\right) \cdot w\big(\gamma_{h_{m,m+1}}(a_m, a_{m+1})\big)\\
&= \prod_{1\le i\le m+1} x_{k_i} \prod_{i\in\{1,\dots,m\}} z_{h_{i,m+1}}.
\end{aligned}
$$

Referring back to (5.23) and letting $\psi' = \bigwedge_{1\le i\le m+1} \beta_{k_i}(a_i)$,

$$
\begin{aligned}
w&\left(\Theta_m(a_1,\dots,a_m) \wedge \beta_{k_{m+1}}(a_{m+1}) \wedge \bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\right)\\
&= w\big(\Theta_m \wedge \beta_{k_{m+1}}(a_{m+1}) \,|\, \psi'\big) \cdot w\left(\bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1}) \,\middle|\, \psi'\right) w(\psi')\\
&= w\big(\Theta_m \wedge \beta_{k_{m+1}}(a_{m+1})\big) \cdot \frac{w\Big(\bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\Big)}{w(\psi')}\\
&= w(\Theta_m) \cdot w(\beta_{k_{m+1}}(a_{m+1})) \cdot \frac{w\Big(\bigwedge_{i\in\{1,\dots,m\}} \gamma_{h_{i,m+1}}(a_i, a_{m+1})\Big)}{w(\psi')}\\
&= \left(\prod_{1\le i\le m} x_{k_i} \prod_{1\le i<t\le m} z_{h_{i,t}}\right) \cdot x_{k_{m+1}} \cdot \prod_{i\in\{1,\dots,m\}} z_{h_{i,m+1}}
\end{aligned}
$$

$$= \prod_{1 \le i \le m+1} x_{k_i} \prod_{1 \le i < t \le m+1} z_{h_{i,t}}.$$

It therefore follows by induction that (5.22) holds for every $m \in \mathbb{N}^+$ and since a probability function is defined by its values on state descriptions, $w = w_{\vec{Y}}$, which completes the proof. □

5.5 A Representation Theorem for BEx

We showed in Proposition 5.6 that the probability functions $w_{\vec{Y}}$ satisfy BEx. We now prove that the functions satisfying BEx are exactly the convex combinations of the $w_{\vec{Y}}$ functions in the following sense.

Theorem 5.9. *Let w be a probability function for a binary language L satisfying BEx. Then there exists a (normalised, σ-additive) measure μ on the Borel subsets of $\mathbb{D}_L$ such that for any $\theta \in SL$,*

$$w(\theta) = \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\ d\mu(\vec{Y}). \tag{5.24}$$

Conversely, for a given measure μ on the Borel subsets of $\mathbb{D}_L$, the function defined by (5.24) is a probability function on SL satisfying BEx.

Proof. Let w be a probability function for L satisfying BEx. It suffices to prove (5.24) for state descriptions, the rest follows, for instance, as in Corollary 9.2 of [36]. The proof is based on the fact that for a state description $\Theta(b_1, \dots, b_m)$ and $u > m$

$$w(\Theta(b_1, \dots, b_m)) = \sum_{\Psi(b_1,\dots,b_m,b_{m+1},\dots,b_u) \models \Theta(b_1,\dots,b_m)} w(\Psi(b_1, \dots, b_m, b_{m+1}, \dots, b_u)), \tag{5.25}$$

and it proceeds via grouping state descriptions for u individuals according to their extended signature and counting their numbers.

Let $t_1, \dots, t_n \in \mathbb{N}$, $t_1 + t_2 + \dots + t_n = t$. We define

$$\binom{t}{\{t_i : i \in \{1, \dots, n\}\}} = \binom{t}{t_1, t_2, \dots, t_n} = \frac{t!}{t_1! t_2! \dots t_n!}.$$

Other expressions using this notation are to be interpreted similarly.

Let $u \in \mathbb{N}^+$ and let $\vec{u}\vec{t} = \langle u_1, \dots, u_{2^q}\,;\, t_1, \dots, t_p \rangle$ be an extended signature on u. Firstly, we wish to count the number of *all* state descriptions with this extended signature. Thinking about state descriptions in terms of $u \times u$ matrices as in (5.5), this involves placing, on the diagonal, the number 1 u_1 times, the number 2 u_2 times and so on. We are thus creating $u_k u_c$ many spaces (when $k \neq c$) or $\frac{u_k(u_k-1)}{2}$ many spaces in which to place atoms from the classes Γ_j, $j \in A(k,c)$ $(k \neq c)$ or $j \in A(k,k)$ respectively. Once a place for an atom from a given Γ_j is chosen, no freedom remains over which atom from this class it is when $k \neq c$ or when $k = c$ and $e = d$ (that is, when $s_j = 1$). When $k = c$ and $e \neq d$ (i.e., when $s_j = 2$), either one of the two atoms from this class can be chosen to fill the place.

It follows that the number of state descriptions with extended signature[10] $\vec{u}\vec{t}$, denoted by $\mathcal{N}(\emptyset, \vec{u}\vec{t})$, is

$$\binom{u}{u_1, \dots, u_{2^q}} \prod_{1 \leq k < c \leq 2^q} \binom{u_k u_c}{\{t_j : j \in A(k,c)\}} \\ \times \prod_{1 \leq k \leq 2^q} \left(\binom{\frac{u_k(u_k-1)}{2}}{\{t_j : j \in A(k,k)\}} \prod_{j \in A(k,k)} s_j^{t_j} \right). \quad (5.26)$$

Now let $\vec{m}\vec{n}$ be an extended signature, $m < u$ and let $\Theta(b_1, \dots, b_m)$ be a state description with this signature. Arguing similarly to above, we find that the number of state descriptions with signature $\vec{u}\vec{t}$ extending $\Theta(b_1, \dots, b_m)$, denoted by $\mathcal{N}(\vec{m}\vec{n}, \vec{u}\vec{t})$, is

$$\binom{u-m}{u_1 - m_1, \dots, u_{2^q} - m_{2^q}} \prod_{1 \leq k < c \leq 2^q} \binom{u_k u_c - m_k m_c}{\{t_j - n_j : j \in A(k,c)\}} \\ \times \prod_{1 \leq k \leq 2^q} \left(\binom{\frac{u_k(u_k-1)}{2} - \frac{m_k(m_k-1)}{2}}{\{t_j - n_j : j \in A(k,k)\}} \prod_{j \in A(k,k)} s_j^{(t_j - n_j)} \right). \quad (5.27)$$

We make the convention that our multinomial expression is 0 if any of the terms are negative[11]. Note that the number calculated in (5.27) depends only on the signature $\vec{m}\vec{n}$ and not on the particular choice of $\Theta(b_1, \dots, b_m)$, since extending any state

[10] Throughout this proof, we will often refer to extended signatures simply as signatures in attempt to reduce cumbersome terminology.

[11] This is justified, since a state description with signature $\vec{u}\vec{t}$ can extend a state description with signature $\vec{m}\vec{n}$ only if $m < u$, $m_k \leq u_k$ for every k and $n_j \leq t_j$ for every j.

description with signature $\vec{m}\vec{n}$ to one with signature $\vec{u}\vec{t}$ involves making the same decisions.

We shall write $w(\vec{m}\vec{n})$ for $w(\Theta(b_1,\ldots,b_m))$; by BEx this is unambiguous. Let $Sign(u)$ denote the set containing all extended signatures $\vec{u}\vec{t}$ on u. That is,

$$Sign(u) = \left\{ \vec{u}\vec{t} : \sum_{k=1}^{2^q} u_k = u, \text{ for } k \neq c \sum_{j\in A(k,c)} t_j = u_k u_c, \sum_{j\in A(k,k)} t_j = \frac{u_k(u_k-1)}{2} \right\}$$

where $1 \leq k, c \leq 2^q$. From (5.25)

$$1 = w(\top) = \sum_{\vec{u}\vec{t}\in Sign(u)} \mathcal{N}(\emptyset, \vec{u}\vec{t})\, w(\vec{u}\vec{t}), \tag{5.28}$$

$$w(\vec{m}\vec{n}) = \sum_{\vec{u}\vec{t}\in Sign(u)} \mathcal{N}(\vec{m}\vec{n}, \vec{u}\vec{t})\; w(\vec{u}\vec{t}), \tag{5.29}$$

and hence

$$w(\vec{m}\vec{n}) = \sum_{\vec{u}\vec{t}\in Sign(u)} \frac{\mathcal{N}(\vec{m}\vec{n}, \vec{u}\vec{t})}{\mathcal{N}(\emptyset, \vec{u}\vec{t})} \mathcal{N}(\emptyset, \vec{u}\vec{t})\; w(\vec{u}\vec{t}). \tag{5.30}$$

We shall show that

$$\left| \left(\frac{\mathcal{N}(\vec{m}\vec{n}, \vec{u}\vec{t})}{\mathcal{N}(\emptyset, \vec{u}\vec{t})} \right) - \left(\prod_{1\leq k\leq 2^q} \left(\frac{u_k}{u}\right)^{m_k} \prod_{1\leq k<c\leq 2^q} \left(\prod_{j\in A(k,c)} \left(\frac{t_j}{u_k u_c} \right)^{n_j} \right) \prod_{1\leq k\leq 2^q} \left(\prod_{j\in A(k,k)} \left(\frac{t_j s_j^{-1}}{\left(\frac{u_k(u_k-1)}{2}\right)} \right)^{n_j} \right) \right) \right| \tag{5.31}$$

is of the order $\mathcal{O}\left(\sqrt{u}^{-1}\right)$ (independently of $u_1,\ldots,u_{2^q},t_1,\ldots,t_p$), so that as $u \to \infty$, (5.31) tends to 0. We make a convention that if some $u_k = 0$ or some $t_j = 0$ then terms involving these are missing from the product above[12].

First, let $m_k \leq u_k$ and $n_j \leq t_j$ for every j,k, so that none of the terms in (5.27) are negative. The term $\left(\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})}\right)$ in (5.31) can be written as

$$\frac{u_1\,(u_1-1)\cdots(u_1-(m_1-1))\cdots u_{2^q}\,(u_{2^q}-1)\cdots(u_{2^q}-(m_{2^q}-1))}{u\,(u-1)\cdots(u-(m-1))}$$

$$\times \prod_{1\leq k<c\leq 2^q} \left(\frac{\prod_{j\in A(k,c)} t_j\,(t_j-1)\cdots(t_j-(n_j-1))}{(u_k u_c)\,(u_k u_c-1)\cdots(u_k u_c-(m_k m_c-1))} \right)$$

[12]Note that this is valid, since factors involving $u_k = 0$ and $t_j = 0$ cancel out from the binomials in $\left(\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})}\right)$, so the convention means the same factors are missing from both terms in (5.31).

$$\times \prod_{1\le k\le 2^q} \left(\frac{\prod_{j\in A(k,k)} t_j\,(t_j-1)\cdots(t_j-(n_j-1))}{\left(\frac{u_k(u_k-1)}{2}\right)\left(\frac{u_k(u_k-1)}{2}-1\right)\cdots\left(\frac{u_k(u_k-1)}{2}-\left(\frac{m_k(m_k-1)}{2}-1\right)\right)} \times \prod_{j\in A(k,k)} s_j^{-n_j} \right),$$

which (using our convention about zero terms) can in turn be written as[13]

$$\left(\prod_{1\le k\le 2^q} \left(\frac{u_k}{u}\right)^{m_k} \prod_{1\le k<c\le 2^q} \left(\prod_{j\in A(k,c)} \left(\frac{t_j}{u_k u_c}\right)^{n_j} \right) \prod_{1\le k\le 2^q} \left(\prod_{j\in A(k,k)} \left(\frac{t_j s_j^{-1}}{\left(\frac{u_k(u_k-1)}{2}\right)} \right)^{n_j} \right) \right) \tag{5.32}$$

$$\times \frac{\prod_{1\le k\le 2^q}\prod_{0\le i\le m_k-1}\left(1-i\,u_k^{-1}\right)}{\prod_{0\le l\le m-1}\left(1-l\,u^{-1}\right)} \tag{5.33}$$

$$\times \prod_{1\le k<c\le 2^q} \left(\frac{\prod_{j\in A(k,c)}\prod_{0\le i\le n_j-1}\left(1-i\,t_j^{-1}\right)}{\prod_{0\le l\le m_k m_c-1}\left(1-l\,(u_k u_c)^{-1}\right)} \right) \tag{5.34}$$

$$\times \prod_{1\le k\le 2^q} \left(\frac{\prod_{j\in A(k,k)}\prod_{0\le i\le n_j-1}\left(1-i\,t_j^{-1}\right)}{\prod_{0\le l\le (m_k(m_k-1)/2)-1}\left(1-l\left(\frac{u_k(u_k-1)}{2}\right)^{-1}\right)} \right). \tag{5.35}$$

Let P stand for the product of (5.33), (5.34) and (5.35).

We observe that P is bounded by a constant independent of u, the u_k and the t_j:

$$(5.33) < \left(\frac{1}{1-(m-1)\,m^{-1}}\right)^m = m^m$$

since (5.33) can be split into the product of m fractions, each numerator is at most 1, and the least denominator is $(1-(m-1)u^{-1})$ which is greater than $(1-(m-1)m^{-1})$. Similarly for (5.34) and (5.35) we have

$$(5.34) < \prod_{1\le k<c\le 2^q} \left(\frac{1}{1-(m_k m_c-1)(m_k m_c)^{-1}}\right)^{m_k m_c} = \prod_{1\le k<c\le 2^q} (m_k m_c)^{m_k m_c},$$

and

$$(5.35) < \prod_{1\le k\le 2^q} \left(\frac{m_k(m_k-1)}{2}\right)^{\frac{m_k(m_k-1)}{2}}.$$

Furthermore, we need only to consider those k where $m_k > 0$ in the limit of (5.31) since otherwise $n_j = 0$ for $j \in A(k,c)$ and factors involving corresponding u_k, t_j cancel out from $\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})}$, and they are all 1 in the product which is being subtracted.

[13] by splitting each of the three products into a term in (5.32) times one of (5.33), (5.34), (5.35).

We shall prove the claim about (5.31) by cases. Consider first the case that for some k with $m_k > 0$ we have $u_k \leq \sqrt{u}$. Then

$$\prod_{1\leq k\leq 2^q}\left(\frac{u_k}{u}\right)^{m_k} \leq (\sqrt{u})^{-1},$$

each of the other products in (5.32) is at most 1, so (5.31) $= |\,(5.32)\cdot(1-P)\,| = \mathcal{O}(\sqrt{u}^{-1})$. A similar argument works if $u_k > \sqrt{u}$ for every k with $m_k > 0$ but for some j we have $n_j > 0$ and $t_j \leq \sqrt{u}$, since then if $j \in A(k,c)$,

$$\prod_{1\leq k<c\leq 2^q}\left(\prod_{j\in A(k,c)}\left(\frac{t_j}{u_k u_c}\right)^{n_j}\right) \leq (\sqrt{u})^{-1}$$

and if $j \in A(k,k)$,

$$\prod_{1\leq k\leq 2^q}\left(\prod_{j\in A(k,k)}\left(\frac{t_j s_j^{-1}}{\left(\frac{u_k(u_k-1)}{2}\right)}\right)^{n_j}\right) \leq 2\,(\sqrt{u}-1)^{-1}$$

which is $\mathcal{O}(\sqrt{u}^{-1})$, the other terms in (5.32) are at most 1, so again (5.31) $= \mathcal{O}(\sqrt{u}^{-1})$.

The second case is when for every k such that $m_k > 0$, $u_k > \sqrt{u}$ and for every j with $n_j > 0$, $t_j > \sqrt{u}$. In this case, P is close to 1. To see this, note that (5.33) can be written as a product of m fractions of the form $\frac{1-\alpha u_k^{-1}}{1-\beta u^{-1}}$, $\alpha, \beta \in \{1,\dots,m\}$ and that the distance of each fraction from 1 is

$$\left|\frac{1-\alpha u_k{}^{-1}}{1-\beta u^{-1}} - 1\right| = \left|\frac{\beta u^{-1} - \alpha u_k^{-1}}{1-\beta u^{-1}}\right| < 2\left(\beta u^{-1} + \alpha u_k{}^{-1}\right) < 2\sqrt{u}^{-1}(\alpha+\beta) \leq 4m\sqrt{u}^{-1},$$

where the inequalities hold since when u is very large, $\beta \leq m < \frac{u}{2}$ and so $1-\beta u^{-1} > \frac{1}{2}$, because $u^{-1}, u_k{}^{-1} < \sqrt{u}^{-1}$ and since $\alpha, \beta \leq m$, respectively. Hence (5.33) is

$$\left(1+\mathcal{O}\left(\sqrt{u}^{-1}\right)\right)^m = 1+\mathcal{O}\left(\sqrt{u}^{-1}\right).$$

A similar argument works for the other two products, (5.34) and (5.35), since for very large u, $m_k m_c < \frac{u_k u_c}{2}$ and $\frac{m_k(m_k-1)}{2} < \frac{u_k(u_k-1)}{4}$, so P is $\left(1+\mathcal{O}\left(\sqrt{u}^{-1}\right)\right)^3 = 1+\mathcal{O}\left(\sqrt{u}^{-1}\right)$. Since each term in (5.32) is bounded by 1, (5.32) is bounded by 1. It follows that (5.31) $= |\,(5.32)\cdot(1-P)\,|$ is again of order $\mathcal{O}\left(\sqrt{u}^{-1}\right)$.

Now suppose $u_k < m_k$ for some k (the case when $u_k > m_k$ for every k but some j is such that $t_j < n_j$ is similar). Note that then $\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})} = 0$ since in that case no state description with signature $\vec{u}\vec{t}$ extending a state description with signature $\vec{m}\vec{n}$

can exist. In addition, $m_k > 0$ and $u_k < \sqrt{u}$ since u is very large and m_k is fixed, so arguing as above (5.32) would be of order $\mathcal{O}(\sqrt{u}^{-1})$ and consequently so would (5.31), which exhausts all cases.

Define $\vec{Y}_{\vec{u}\vec{t}}$ by

$$x_k = \frac{u_k}{u}, \qquad y_j = \begin{cases} \frac{t_j}{u_k u_c} & \text{for } j \in A(k,c),\ u_k, u_c \neq 0,\ k < c, \\ \frac{t_j s_j{}^{-1}}{\binom{\frac{u_k(u_k-1)}{2}}{}} & \text{for } j \in A(k,k),\ u_k \neq 0, 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.36}$$

At this point, we can complete the proof using classical techniques, or we can employ methods from Nonstandard Analysis, particularly Loeb Measure Theory [10, 26]. We present both, beginning with the classical proof.

From (5.28), let μ_u be the normalised discrete measure on $\mathbb{D}_L$ which puts measure

$$\mathcal{N}(\emptyset, \vec{u}\vec{t})\, w(\vec{u}\vec{t})$$

on the point $\vec{Y}_{\vec{u}\vec{t}}$. Using (5.28), (5.30) and the fact that (5.32) gets uniformly close to $\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})}$ as $u \to \infty$ gives that $w(\vec{m}\vec{n})$ equals the limit as $u \to \infty$ of

$$\sum_{\vec{u}\vec{t} \in Sign(u)} \left(\prod_{1 \le k \le 2^q} \left(\frac{u_k}{u}\right)^{m_k} \prod_{1 \le k < c \le 2^q} \left(\prod_{j \in A(k,c)} \frac{t_j}{u_k u_c} \right)^{n_j} \right.$$
$$\left. \times \prod_{1 \le k \le 2^q} \left(\prod_{j \in A(k,k)} \left(\frac{t_j s_j^{-1}}{\frac{u_k(u_k-1)}{2}} \right)^{n_j} \right) \right) \mu_u\,(\vec{Y}_{\vec{u}\vec{t}}). \tag{5.37}$$

So

$$w(\vec{m}\vec{n}) = \lim_{u \to \infty} \int_{\mathbb{D}_L} \left(\prod_{1 \le k \le 2^q} x_k^{m_k} \cdot \prod_{1 \le k < c \le 2^q} \prod_{j \in A(k,c)} y_j^{n_j} \cdot \prod_{1 \le k \le 2^q} \prod_{j \in A(k,k)} y_j^{n_j} \right) d\mu_u(\vec{Y}) \tag{5.38}$$

where x_k, y_j are as in (5.36).

Following the same argument as in the proof of [36, Theorem 9.1], using Prohorov's Theorem [1, Theorem 5.1] we have that since $\mathbb{D}_L$ is compact, the μ_u have a subsequence μ_{u_i} which is weakly convergent to a countably additive measure μ. So for any continuous function f in variables $x_1, \ldots, x_{2^q}, y_1, \ldots, y_p$, we have that

$$\lim_{u\to\infty}\int_{\mathbb{D}_L} f \;\; d\mu_{u_i}(\vec{Y}) = \int_{\mathbb{D}_L} f \;\; d\mu(\vec{Y}).$$

Now using (5.38)

$$\begin{aligned} w(\vec{m}\vec{n}) &= \int_{\mathbb{D}_L}\left(\prod_{1\le k\le 2^q} {x_k}^{m_k}\cdot\prod_{1\le k<c\le 2^q}\prod_{j\in A(k,c)} y_j^{n_j}\cdot\prod_{1\le k\le 2^q}\prod_{j\in A(k,k)} y_j^{n_j}\right)\; d\mu(\vec{Y}) \\ &= \int_{\mathbb{D}_L} w_{\vec{Y}}(\vec{m}\vec{n})\;\; d\mu(\vec{Y}). \end{aligned} \tag{5.39}$$

In the opposite direction, a function on SL defined by (5.24) clearly satisfies (P1) and (P2), and by the Lebesgue Dominated Convergence Theorem it also satisfies (P3). So it is a probability function. This function satisfies BEx because all the $w_{\vec{Y}}$ do.

Nonstandard Proof

In what follows, we will write $w_{\vec{u}\vec{t}}$ for $w_{\vec{Y}_{\vec{u}\vec{t}}}$. Note that $w_{\vec{u}\vec{t}}(\vec{m}\vec{n})$ is equal to (5.32).

Let U^* be a nonstandard ω_1-saturated elementary extension of a sufficiently large portion U of the set theoretic universe containing w. As usual, c^* denotes the image in U^* of $c \in U$ where these differ. Working now in U^*, let $u \in \mathbb{N}^*$ be nonstandard. Then (from (5.30)) we still have

$$w^*(\vec{m}\vec{n}) = \sum_{\vec{u}\vec{t}\in Sign^*(u)} \frac{\mathcal{N}^*(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}^*(\emptyset,\vec{u}\vec{t})}\mathcal{N}^*(\emptyset,\vec{u}\vec{t})w^*(\vec{u}\vec{t}). \tag{5.40}$$

Loeb Measure Theory enables us to conclude from (5.40) that for some σ-additive measure μ' on $Sign^*(u)$ we have (for all standard extended signatures $\vec{m}\vec{n}$)

$$w(\vec{m}\vec{n}) = \int_{Sign^*(u)} {}^\circ\left(\frac{\mathcal{N}^*(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}^*(\emptyset,\vec{u}\vec{t})}\right) d\mu'(\vec{u}\vec{t}) \tag{5.41}$$

where $^\circ$ denotes the standard part. Since, in U, (5.31) is $\mathcal{O}\left(\sqrt{u}^{-1}\right)$, this gives

$$w(\vec{m}\vec{n}) = \int_{Sign^*(u)} {}^\circ(w^*_{\vec{u}\vec{t}}(\vec{m}\vec{n}))\; d\mu'(\vec{u}\vec{t})\,. \tag{5.42}$$

Moreover, ${}^\circ\big(w^*_{\vec{u}\vec{t}}(\vec{m}\vec{n})\big)$ equals $w_{({}^\circ(\vec{Y}_{\vec{u}\vec{t}}))}(\vec{m}\vec{n})$. So defining μ on the Borel subsets A of $\mathbb{D}_L$ by

$$\mu(A) = \mu'\Big\{\vec{u}\vec{t} \mid {}^\circ(\vec{Y}_{\vec{u}\vec{t}}) = \langle {}^\circ x_1,\ldots,{}^\circ x_{2^q};{}^\circ y_1,\ldots,{}^\circ y_p\rangle \in A\Big\}$$

where the x_k, y_j are as defined in (5.36), means (5.42) becomes (using, for example, Proposition 1, Chapter 15 of [41])

$$w(\vec{m}\vec{n}) = \int_{\mathbb{D}_L} w_{\vec{Y}}(\vec{m}\vec{n})\, d\mu(\vec{Y}),$$

as required. □

We shall now use the above representation theorem to show that the $w_{\vec{Y}}$ functions, which by Theorem 5.8 are the only probability functions satisfying Ex and BIP, can be characterised alternatively as the only probability functions satisfying BEx and IP. The fact that the $w_{\vec{Y}}$ satisfy BEx and IP follows from Proposition 5.6 and Theorem 5.8, and the other part follows from the following theorem.

Theorem 5.10. *Let w be a probability function on SL satisfying BEx and IP. Then w is equal to $w_{\vec{Y}}$ for some $\vec{Y} \in \mathbb{D}_L$.*

Proof. [14] Let μ be the σ-additive normalised measure guaranteed to exist by Theorem 5.9 such that

$$w = \int_{\mathbb{D}_L} w_{\vec{Y}}\ d\mu(\vec{Y})\,.$$

Let $\theta(b_1, \ldots, b_m) \in QFSL$ and let θ' be the result of replacing each b_i in θ by b_{i+m}. By IP and since $w(\theta) = w(\theta')$ by (B)Ex, we have

$$\begin{aligned}
0 &= 2\left(w(\theta \wedge \theta') - w(\theta) \cdot w(\theta')\right) \\
&= \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta \wedge \theta')\ d\mu(\vec{Y}) + \int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta \wedge \theta')\ d\mu(\vec{Y'}) \\
&\quad -2\left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\ d\mu(\vec{Y})\right) \cdot \left(\int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta')\ d\mu(\vec{Y'})\right) \\
&= \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta) \cdot w_{\vec{Y}}(\theta')\ d\mu(\vec{Y}) + \int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta) \cdot w_{\vec{Y'}}(\theta')\ d\mu(\vec{Y'}) \\
&\quad -2\left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\ d\mu(\vec{Y})\right) \cdot \left(\int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta')\ d\mu(\vec{Y'})\right) \\
&= \int_{\mathbb{D}_L}\int_{\mathbb{D}_L} \left(w_{\vec{Y}}(\theta) - w_{\vec{Y'}}(\theta')\right)^2\ d\mu(\vec{Y})\ d\mu(\vec{Y'}).
\end{aligned}$$

Let $\mathbb{D}_L^{(1)} = \mathbb{D}_L$,

$$A_{11} = \left\{\vec{Y} \in \mathbb{D}_L\ :\ w_{\vec{Y}}(\theta) \in \left[0, \frac{1}{3}\right)\right\},$$

[14] This is based on the method of the proof of [36, Theorem 20.6].

$$A_{12} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) \in \left[\frac{1}{3}, \frac{2}{3}\right) \right\},$$
$$A_{13} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) \in \left[\frac{2}{3}, 1\right] \right\}.$$

Let $\mu(A_{11}) = a_{11}$, $\mu(A_{12}) = a_{12}$ and $\mu(A_{13}) = a_{13}$, so in particular $\mathbb{D}_L^{(1)} = A_{11} \cup A_{12} \cup A_{13}$, $\mu(\mathbb{D}_L^{(1)}) = a_{11} + a_{12} + a_{13} = 1$. Then

$$\begin{aligned} 0 &= \int_{\mathbb{D}_L} \int_{\mathbb{D}_L} \left(w_{\vec{Y}}(\theta) - w_{\vec{Y'}}(\theta') \right)^2 \, d\mu(\vec{Y}) \, d\mu(\vec{Y'}) \\ &\geq \int_{A_{11}} \int_{A_{13}} \left(w_{\vec{Y}}(\theta) - w_{\vec{Y'}}(\theta') \right)^2 \, d\mu(\vec{Y}) \, d\mu(\vec{Y'}) \\ &\geq \left(\frac{1}{3}\right)^2 a_{11} a_{13} \end{aligned}$$

since we are integrating over a positive function so integrating over a subset of $\mathbb{D}_L$ can at most not decrease the integral, and since for $\vec{Y'} \in A_{11}$ and $\vec{Y} \in A_{13}$, $\left(w_{\vec{Y}}(\theta) - w_{\vec{Y'}}(\theta') \right)^2 \geq \left(\frac{1}{3}\right)^2$.

So one of a_{11}, a_{13} must be 0. Without loss of generality suppose $a_{13} = 0$, then $a_{11} + a_{12} = 1$. Let

$$\mathbb{D}_L^{(2)} = \left\{ \vec{Y} \,:\, w_{\vec{Y}}(\theta) \in \left[0, \frac{2}{3}\right) \right\}$$

so $\mu(\mathbb{D}_L^{(2)}) = 1$, and let

$$A_{21} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) \in \left[0, \frac{2}{9}\right) \right\},$$
$$A_{22} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) \in \left[\frac{2}{9}, \frac{4}{9}\right) \right\},$$
$$A_{23} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) \in \left[\frac{4}{9}, \frac{2}{3}\right) \right\}.$$

Let $\mu(A_{21}) = a_{21}$, $\mu(A_{22}) = a_{22}$, $\mu(A_{23}) = a_{23}$. Following the same argument as above,

$$0 \geq \left(\frac{2}{9}\right)^2 a_{21} a_{23},$$

so one of a_{21}, a_{23} must be zero. We proceed by picking one of these to be non-zero and splitting the remaining interval into three again.

Repeating this process infinitely many times, we obtain a sequence $\mathbb{D}_L^{(n)}$, $n \in \mathbb{N}^+$, where

$$\mathbb{D}_L^{(n)} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) \in I_n \right\},$$

$I_n \subseteq [0,1]$, $l(I_n) = \left(\frac{2}{3}\right)^{n-1}$ and $\mu(\mathbb{D}_L^{(n)}) = 1$.

Let $B_\theta = \bigcap_{n=1}^{\infty} \mathbb{D}_L^{(n)}$. Then

$$B_\theta = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta) = b \right\}$$

for some $b \in [0,1]$. The complement of B_θ, ${B_\theta}^c$, is $\bigcup_{n=1}^{\infty} \left(\mathbb{D}_L^{(n)}\right)^c$, and $\mu\left(\left(\mathbb{D}_L^{(n)}\right)^c\right) = 0$ for each n since μ is σ-additive. Thus ${B_\theta}^c$ is the union of countably many sets, each with μ measure 0, so $\mu({B_\theta}^c) = 0$ by the σ-additivity of μ, hence $\mu(B_\theta) = 1$. I.e. B_θ is a subset of $\mathbb{D}_L$ with μ measure 1 such that for this particular $\theta \in QFSL$, $w_{\vec{Y}}(\theta)$ as a function of $\vec{Y}$ is constant on B_θ.

Next, we enumerate the countably many quantifier free sentences of L by $\theta_1, \theta_2, \ldots$. Following the above argument, for each sentence θ_i there is an associated

$$B_{\theta_i} = \left\{ \vec{Y} \in \mathbb{D}_L \,:\, w_{\vec{Y}}(\theta_i) = b_i \right\}$$

for some $b_i \in [0,1]$ such that $\mu(B_{\theta_i}) = 1$. Define $B = \bigcap_{i=1}^{\infty} B_{\theta_i}$. Using the same σ-additivity argument as above, $\mu(B^c) = 0$, and so $\mu(B) = 1$. We conclude that B is a subset of $\mathbb{D}_L$ with μ measure 1 such that $w_{\vec{Y}}(\theta)$ as a function of $\vec{Y}$ is constant on B for *every* $\theta \in QFSL$. Therefore, for any $\vec{Y} \in B$ we must have that w and $w_{\vec{Y}}$ are equal for quantifier free sentences and hence for all sentences, which proves the theorem. $\square$

In fact, as long as none of the x_k in $\vec{Y}$ are zero, B as defined above contains a single $\vec{Y}$ from $\mathbb{D}_L$:

Proposition 5.11. *Let $\vec{Y}, \vec{Y'} \in \mathbb{D}_L$ be such that $x_1, \ldots, x_{2^q}, x_1', \ldots, x_{2^q}' > 0$ and $\vec{Y} \neq \vec{Y'}$. Then there exists Θ a state description of L such that*

$$w_{\vec{Y}}(\Theta) \neq w_{\vec{Y'}}(\Theta).$$

Proof. Suppose first that for some $s \in \{1, \ldots, 2^q\}$, we have[15] $x_s \neq x_s'$. Let $\Theta(a_1) = \beta_s(a_1)$. Then $w_{\vec{Y}}(\Theta(a_1)) = x_s$, $w_{\vec{Y'}}(\Theta(a_1)) = x_s'$ and these are not equal. So suppose that $x_s = x_s'$ for every $s \in \{1, \ldots, 2^q\}$, but $y_j \neq y_j'$ for some $j \in \{1, \ldots, p\}$. In this case let $\Theta(a_1, a_2) = \beta_k(a_1) \wedge \beta_c(a_2) \wedge \gamma_h(a_1, a_2)$, with $\gamma_h \in \Gamma_j$. Then $w_{\vec{Y}}(\Theta(a_1, a_2)) = x_k\, x_c\, y_j$ and $w_{\vec{Y'}}(\Theta(a_1, a_2)) = x_k\, x_c\, y_j'$ and these are not equal whenever $x_k, x_c > 0$. $\square$

[15] In fact, in this case at least two entries in $x_1, \ldots, x_{2^q}$ and $x_1', \ldots, x_{2^q}'$ must differ since the x_k add up to 1.

We have characterised the $w_{\vec{Y}}$ as the only probability functions satisfying BIP and Ex, and the only functions satisfying IP and BEx. In contrast to the $w_{\vec{x}}$ in the unary however, the $w_{\vec{Y}}$ are *not* the only functions satisfying IP and Ex.

Proposition 5.12. *There exists a probability function w such that w satisfies Ex and IP and $w \neq w_{\vec{Y}}$ for any $\vec{Y} \in \mathbb{D}_L$.*

Proof. We use results and notation from [36, Chapter 25]. Let L be a language containing a single binary relation symbol R. Let w be a probability function on SL satisfying Ex and let U^* be a nonstandard universe as defined on page 95. Let $n \in \mathbb{N}$, let $\nu \in \mathbb{N}^*$ be nonstandard, and let $\Theta(a_1, \ldots, a_n)$ be a state description of L (so it is a state description also in U^*) and $\Psi(a_1, \ldots, a_\nu)$ a state description in U^*.

We pick in U^*, uniformly at random and with replacement, $a_{h_1}, a_{h_2}, \ldots, a_{h_n}$ from $\{a_1, a_2, \ldots, a_\nu\}$ and define $w^\Psi(\Theta(a_1, \ldots, a_n))$ as the probability that

$$\Psi(a_1, \ldots, a_\nu) \models \Theta(a_{h_1}, \ldots, a_{h_n}).$$

From [36, Chapter 25] we have that the ${}^\circ w^\Psi$, the standard part of w^Ψ, are (standard) probability functions on SL which satisfy Ex and IP and moreover, a probability function w satisfies Ex and IP just if $w = {}^\circ w^\Psi$ for some state description $\Psi(a_1, \ldots, a_\nu)$ in U^*.

The $w_{\vec{Y}}$ satisfy BEx, so we shall show that there exists a probability function ${}^\circ w^\Psi$ that gives state descriptions with the same signature different probabilities and so cannot be equal to a $w_{\vec{Y}}$ for any $\vec{Y} \in \mathbb{D}_L$.

Let $\Theta(a_1, a_2, a_3, a_4)$, $\Phi(a_1, a_2, a_3, a_4)$ be state descriptions of L represented respectively by the following matrices:

$$\begin{matrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \qquad\qquad \begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{matrix}.$$

Let $\Psi(a_1, \ldots, a_\nu)$ be a state description in U^* such that $\Psi \models R(a_i, a_i)$ for each $1 \leq i \leq \nu$, $\Psi \models R(a_i, a_t)$ for each $\langle i, t \rangle$ with $1 \leq i, t \leq \frac{\nu}{2}$, and $\Psi \models \neg R(a_i, a_t)$ otherwise. Then Θ

and Φ have the same signature, but $^\circ w^\Psi(\Theta) > 0$ and $^\circ w^\Psi(\Phi) = 0$. This is because there is no way of choosing $a_{h_1}, a_{h_2}, a_{h_3}, a_{h_4}$ from $a_1, \ldots, a_\nu$ so that $\Psi[a_{h_1}, a_{h_2}, a_{h_3}, a_{h_4}] \equiv \Phi(a_{h_1}, a_{h_2}, a_{h_3}, a_{h_4})$, but $\Psi[a_{h_1}, a_{h_2}, a_{h_3}, a_{h_4}] \equiv \Theta(a_{h_1}, a_{h_2}, a_{h_3}, a_{h_4})$ whenever $1 \leq h_1, h_2, h_3 \leq \frac{\nu}{2}$ and $\frac{\nu}{2} < h_4 \leq \nu$ (so in fact $^\circ w^\Psi(\Theta) = \frac{1}{16}$). $\square$

5.6 Binary Instantial Relevance

In this section we consider how the idea of instantial relevance might be captured in our atom-based binary context. Assuming that the available evidence is in the form of a partial state description, the evidence may be extended to another partial state description either by adding unary information about a new individual, or by adding a binary atom instantiated by a pair of individuals each of which may or may not be new. In each of these cases, if we have already learnt (and added to the evidence) the same information about another individual or pair of individuals, it should enhance our probability that this information will be learnt about the given individual or pair of individuals too.

Adding unary information about a single constant does not involve any intricacies, and instantial relevance amounts to requiring that for a partial state description $\Delta(a_1, \ldots, a_m)$ and any β_k,

$$w(\beta_k(a_{m+2}) \mid \Delta) \;\leq\; w(\beta_k(a_{m+2}) \mid \beta_k(a_{m+1}) \wedge \Delta). \tag{5.43}$$

Adding an atom instantiated by some constants b_1, b_2 is more complicated, since such sentences are already determined to some degree by Δ when one or both of b_1, b_2 are amongst the $a_1, \ldots, a_m$. More precisely, assume that

$$\gamma_h(b_1, b_2) \wedge \Delta(a_1, \ldots, a_m)$$

is consistent and that $\beta_k(x) \wedge \beta_c(y)$ is the unary trace of $\gamma_h(x, y)$. Then $\Delta(a_1, \ldots, a_m)$ may already imply $\gamma_h(b_1, b_2)$, or imply only $\beta_k(b_1) \wedge \beta_c(b_2)$, or only $\beta_k(b_1)$, or only $\beta_c(b_2)$, or none of these. According to which of these holds, we define the *Extra* in $\gamma_h(b_1, b_2)$ over $\Delta(a_1, \ldots, a_m)$ to be, in order,

$$\emptyset, \quad \{\{1,2\}\}, \quad \{\{1,2\},\{2\}\}, \quad \{\{1,2\},\{1\}\}, \quad \{\{1,2\},\{1\},\{2\}\}$$

respectively. Naturally, conditional probabilities of instantiated atoms given partial state descriptions should only be compared if the Extra in them over the evidence is the same.

Binary Principle of Instantial Relevance

Let $\Delta(a_1,\dots,a_m)$ be a partial state description. Then (5.43) holds for any β_k. Furthermore, if γ_h is an atom and b_1,b_2,b_1',b_2' are constants such that $\Delta\wedge\gamma_h(b_1,b_2)\wedge\gamma_h(b_1',b_2')$ is consistent and the Extras in $\gamma_h(b_1,b_2)$ over $\Delta\wedge\gamma(b_1',b_2')$, in $\gamma_h(b_1,b_2)$ over Δ and in $\gamma_h(b_1',b_2')$ over Δ are all the same then

$$w(\gamma_h(b_1,b_2)\,|\,\Delta)\leq w(\gamma_h(b_1,b_2)\,|\,\gamma_h(b_1',b_2')\wedge\Delta). \tag{5.44}$$

Theorem 5.13. *Let w be a probability function on SL satisfying BEx. Then w satisfies the Binary Principle of Instantial Relevance.*

Proof. Firstly, note that every $w_{\vec{Y}}$ satisfies (5.43) and (5.44) with equality by the definition of these functions. To see this, let

$$\Delta(a_1,\dots,a_m)=\bigwedge_{1\leq i\leq m}\beta_{k_i}(a_i)\wedge\bigwedge_{\substack{\{a_i,a_t\}\in C\\ i<t}}\gamma_{h_{i,t}}(a_i,a_t).$$

Using the fact that the $w_{\vec{Y}}$ satisfy (B)IP and that the following sentences are all partial state descriptions,

$$w_{\vec{Y}}(\Delta\wedge\beta_k(a_{m+2}))=w_{\vec{Y}}(\Delta)\cdot x_k\,,$$

$$w_{\vec{Y}}(\Delta\wedge\beta_k(a_{m+2})\wedge\beta_k(a_{m+1}))=w_{\vec{Y}}(\Delta)\cdot x_k^2\,.$$

and, for example, when the above Extra is $\{\{1,2\},\{2\}\}$ and the unary trace of $\gamma_h(x,y)$ is $\beta_k(x)\wedge\beta_c(y)$, then

$$\Delta\wedge\gamma_h(b_1,b_2)=\bigwedge_{1\leq i\leq m}\beta_{k_i}(a_i)\wedge\bigwedge_{\substack{\{a_i,a_t\}\in C\\ i<t}}\gamma_{h_{i,t}}(a_i,a_t)\wedge\beta_c(b_2)\wedge\gamma_h(b_1,b_2)$$

$$\begin{aligned}\Delta\wedge\gamma_h(b_1,b_2)\wedge\gamma_h(b_1',b_2')=&\bigwedge_{1\leq i\leq m}\beta_{k_i}(a_i)\wedge\bigwedge_{\substack{\{a_i,a_t\}\in C\\ i<t}}\gamma_{h_{i,t}}(a_i,a_t)\\&\wedge\beta_c(b_2)\wedge\beta_c(b_2')\wedge\gamma_h(b_1,b_2)\wedge\gamma_h(b_1',b_2')\end{aligned}$$

and

$$w(\Delta \wedge \gamma_h(b_1, b_2)) = w(\Delta \wedge \gamma_h(b_1', b_2')) = w(\Delta) \cdot x_c \cdot z_h\,,$$

$$w(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma(b_1', b_2')) = w(\Delta) \cdot x_c^2 \cdot z_h^2\,.$$

Similarly, when the Extra is $\{\{1,2\}\}$, $\{\{1,2\},\{1\}\}$ or $\{\{1,2\},\{1\},\{2\}\}$, then

$$w_{\vec{Y}}(\Delta \wedge \gamma_h(b_1, b_2)) = w_{\vec{Y}}(\Delta \wedge \gamma_h(b_1', b_2')) = w_{\vec{Y}}(\Delta) \cdot f(\vec{Y})$$

$$w_{\vec{Y}}(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b_1', b_2')) = w_{\vec{Y}}(\Delta) \cdot (f(\vec{Y}))^2,$$

with $f(\vec{Y})$ respectively (z_h), $(z_h \cdot x_k)$ and $(z_h \cdot x_k \cdot x_c)$. When the Extra is $\emptyset$, (5.44) holds trivially for any w, since in that case Δ, $\Delta \wedge \gamma_h(b_1, b_2)$, $\Delta \wedge \gamma_h(b_1', b_2')$ and $\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b_1', b_2')$ are all logically equivalent.

By Theorem 5.9, since w satisfies BEx, w is an integral of the $w_{\vec{Y}}$. Let μ be the corresponding measure. Then (5.44) (and similarly[16] (5.43)), can be expressed as

$$\frac{w(\Delta \wedge \gamma_h(b_1, b_2))}{w(\Delta)} \leq \frac{w(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b_1', b_2'))}{w(\Delta \wedge \gamma_h(b_1', b_2'))}$$

$$(w(\Delta \wedge \gamma_h(b_1, b_2)))^2 \leq w(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b_1', b_2')) \cdot w(\Delta)$$

$$\begin{aligned} \left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\Delta \wedge \gamma_h(b_1, b_2))\, d\mu(\vec{Y})\right)^2 &\leq \left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b_1', b_2'))\, d\mu(\vec{Y})\right) \\ &\quad \cdot \left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y})\right) \\ \left(\int_{\mathbb{D}_L} f(\vec{Y}) w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y})\right)^2 &\leq \left(\int_{\mathbb{D}_L} (f(\vec{Y}))^2 w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y})\right) \cdot \left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y})\right). \end{aligned}$$

Let $A = w(\Delta)$, $B = w(\Delta \wedge \gamma_h(b_1, b_2))$ and $C = w(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b_1', b_2'))$. Then the above amounts to $B^2 \leq CA$ or $0 \leq CA - B^2$, and this integral inequality holds for any f, as follows: If $A = 0$, it clearly holds, since $0 \leq B, C \leq A$ by, for example, [36, Proposition 3.1(c)]. If $A > 0$ then $0 \leq CA - B^2 \iff 0 \leq CA^2 - AB^2$, and

$$\begin{aligned} 0 &\leq CA^2 - AB^2 \\ &= CA^2 - 2AB^2 + B^2 A \\ &= \int_{\mathbb{D}_L} A^2\, (f(\vec{Y}))^2\, w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y}) - \int_{\mathbb{D}_L} 2AB\, f(\vec{Y})\, w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y}) + \int_{\mathbb{D}_L} B^2\, w_{\vec{Y}}(\Delta)\, d\mu(\vec{Y}) \end{aligned}$$

[16] Using the same argument, with $\gamma_h(b_1, b_2)$ replaced by $\beta_k(a_{m+2})$, $\gamma_h(b_1', b_2')$ replaced by $\beta_k(a_{m+1})$ and $f(\vec{Y}) = x_k$.

$$= \int_{\mathbb{D}_L} \left(A^2\, (f(\vec{Y}))^2 - 2AB\, f(\vec{Y}) + B^2 \right) w_{\vec{Y}}(\Delta)\; d\mu(\vec{Y})$$
$$= \int_{\mathbb{D}_L} \left(A\, f(\vec{Y}) - B \right)^2 w_{\vec{Y}}(\Delta)\; d\mu(\vec{Y})$$

using that A and B are constants, and this clearly holds since the right hand side is an integral of a non-negative function, thus proving the theorem. □

We remark that the same method yields the following related result:

Theorem 5.14. *Let w be a probability function on SL satisfying BEx. Let $\Delta(a_1, \ldots, a_m)$ be a partial state description. If γ_h is an atom and b_1, b_2, b'_1, b'_2 are constants such that $\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma_h(b'_1, b'_2)$ is consistent and the Extra in $\gamma_h(b'_1, b'_2)$ over Δ is the same as the Extra in $\gamma_h(b_1, b_2)$ over $\Delta \wedge \gamma(b'_1, b'_2)$ then*

$$w(\gamma_h(b'_1, b'_2) \mid \Delta) \leq w(\gamma_h(b_1, b_2) \mid \gamma_h(b'_1, b'_2) \wedge \Delta)\,. \tag{5.45}$$

The main difference[17] between (5.44) and (5.45) is a motivational one. The former is the assertion that the probability of a particular atom occurring is increased (or at least, not decreased) if we have evidence of another pair of individuals instantiating this same atom. When restricted to unary languages, this corresponds to Carnap's formulation of PIR. (5.45) on the other hand, represents the idea that we learn information about pairs of individuals successively, so that once we have learnt that a pair of individuals satisfies a certain atom, this information is added to our evidence. The probability we will then learn that another pair of individuals satisfies this same atom should be at least as much as it was for learning the first pair satisfied this atom.

[17]Note that Theorem 5.14 also makes one less assumption about the required Extras than the statement of the binary PIR.

Chapter 6

Polyadic Signature Exchangeability

We extend our investigation from Chapter 5 to polyadic languages. Specifically, we define the notion of a signature and the Principle of Signature Exchangeability in this more general context. We generalise the $w_{\vec{Y}}$ for polyadic languages and provide a polyadic principle of instantial relevance. Finally, in the last section, we indicate some possible directions for future research to continue this investigation. The results in this chapter appear also in [40].

With the required background and motivation already covered by Section 5.1, we move straight to our results.

6.1 An Atom-based Approach for Polyadic Languages

Let L be an r-ary language with relation symbols $R_1, \ldots, R_q$ of arities $r_1, \ldots, r_q$. So, as usual, the maximum of the r_i is r. Recall that the atoms of L, as defined in Chapter 4 on page 52, are the state formulae for r variables, denoted by

$$\gamma_1(x_1, \ldots, x_r), \ldots, \gamma_N(x_1, \ldots, x_r),$$

and that state descriptions can be expressed as a conjunction of instantiated atoms,

$$\Theta(b_1, \ldots, b_m) = \bigwedge_{\langle i_1, i_2, \ldots, i_r \rangle \in \{1, \ldots, m\}^r} \gamma_{h_{i_1, \ldots, i_r}}(b_{i_1}, \ldots, b_{i_r})$$

which, when $m \geq r$, we can write also as

$$\Theta(b_1, \ldots, b_m) = \bigwedge_{1<i_1<\ldots<i_r\leq m} \gamma_{h_{i_1,\ldots,i_r}}(b_{i_1}, \ldots, b_{i_r}). \tag{6.1}$$

For the purpose of this chapter, we will find it convenient in addition to have a way of referring to blocks smaller than atoms.

Definition 6.1. The *g-atoms* for $g \leq r$ are the state formulae of L for g variables. They are denoted by

$$\gamma_1^g(x_1, \ldots, x_g), \ldots, \gamma_{N_g}^g(x_1, \ldots, x_g).$$

Thus the $\gamma_h^r(x_1, \ldots, x_r)$ are just the atoms $\gamma_h(x_1, \ldots, x_r)$ and $N_r = N$. Clearly, like N, the N_g depend on L. Note that in the binary case there are the $\gamma_h^2 = \gamma_h$ (the binary atoms) and the γ_k^1 (1-atoms), which we referred to as β_k in the previous chapter to avoid superscripts altogether.

As before, any conjunction of atoms is consistent (and hence defines a state description) just when any pair of the $\gamma_{h_{i_1,\ldots,i_r}}(b_{i_1}, \ldots, b_{i_r})$ agree when restricted to the constants they have in common. We will find it useful to make these shared components visible so we write

$$\Theta(b_1, \ldots, b_m) \; = \; \bigwedge_{1\leq s\leq r} \; \bigwedge_{1\leq i_1<\ldots<i_s\leq m} \gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1}, \ldots, b_{i_s})\,. \tag{6.2}$$

This works even when $m < r$. Note that the $\gamma^s_{h_{i_1,\ldots,i_s}}$ in (6.2) are such that

$$\gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1}, \ldots, b_{i_s}) = \Theta[b_{i_1}, \ldots, b_{i_s}].$$

Let $g < r$. The following definition is motivated by the need to isolate the part of a state description in which at most g constants are brought together instantiating a relation. We refer to this part as the *g-ary trace* of the state description. More precisely,

Definition 6.2. The *g-ary trace* of the state description (6.2), denoted by[1]

$$(\Theta \restriction g)(b_1, \ldots, b_m),$$

[1] or sometimes simply by $(\Theta \restriction g)$.

is defined to be

$$\bigwedge_{1\leq s\leq g} \bigwedge_{1\leq i_1<\ldots<i_s\leq m} \gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1},\ldots,b_{i_s}). \tag{6.3}$$

Note that when $g = 1$ this definition agrees with the definition of the unary trace from page 69. Any consistent conjunction of the form (6.3) is called a *g-ary trace for the constants* $b_1,\ldots,b_m$.

Partial state descriptions are composed of instantiated s-atoms in a similar way to state descriptions, but the sentences do not necessarily combine to give a full state description.

Definition 6.3. A *partial state description* for $b_1,\ldots,b_m$ is a sentence of the form

$$\Delta(b_1,\ldots,b_m) = \bigwedge_{1\leq s\leq r} \bigwedge_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s \\ i_1<\ldots<i_s}} \gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1},\ldots,b_{i_s}), \tag{6.4}$$

where C^s is some set of s-element subsets of $\{b_1,\ldots,b_m\}$.

We will assume that (6.4), like (6.2), displays all the instantiated γ^s_h implied by Δ. In other words, we assume that $\bigcup_{s=1}^r C^s$ contains along with any $\{b_{i_1},\ldots,b_{i_s}\}$ also *all* its subsets.

In addition, when writing $\Delta(b_1,\ldots,b_m)$ for a partial state description, we mean that all of $b_1,\ldots,b_m$ actually appear in it, so C^1 contains all singletons $\{b_i\}$ for $i = 1,\ldots,m$. When $r = 2$, this definition agrees with the definition of a partial state description in the binary case (page 70). We remark also that a partial state description (6.4) is a state description just when C^r contains *all* r-element subsets of $\{b_1,\ldots,b_m\}$. Note that any g-ary trace of a state description is a partial state description.

We define the *g-ary trace of a state formula*, and a *partial state formula* analogously to the definitions for state descriptions.

6.2 Polyadic Signatures

As in the binary case, we need to introduce an equivalence between atoms (and more generally, between g-atoms) to capture the fact that g-atoms obtained from each other by permuting the variables represent the same thing.

Accordingly, we define $\gamma_h^g \sim \gamma_k^g$ if there exists a permutation $\sigma \in \mathsf{S}_g$ such that

$$\gamma_h^g(x_1, \ldots, x_g) \equiv \gamma_k^g(x_{\sigma(1)}, \ldots, x_{\sigma(g)}) \tag{6.5}$$

and we denote the equivalence classes of $\sim$ by $\Gamma_1^g, \ldots, \Gamma_{p_g}^g$. When $g = r$ we drop the superscript and write just $\Gamma_1 \ldots, \Gamma_p$, and we write p for p_r. If (6.5) holds, we say that γ_h^g *obtains from* γ_k^g *via* σ. Note that the equivalence classes Γ_j^1 are singletons and $p_1 = N_1 = 2^q$, so they are not necessary and we can work with the γ_k^1 instead, as we did with the β_k in the previous chapter, for $r = 2$.

For $1 < g \leq r$, every Γ_j^g can be split into subclasses, each subclass containing all γ_h^g with the same $(g-1)$-ary trace. Define s_j^g to be the number of elements in these subclasses. This is possible since given g and j, the subclasses of Γ_j^g all have the same number of elements. s_j^g expresses in how many ways the $(g-1)$-ary trace of some/any γ_h^g from Γ_j^g can be extended to a $\gamma_k^g \in \Gamma_j^g$; one of these ways is to γ_h^g itself but there may be other possibilities. In the binary case, we wrote just s_j for s_j^2.

We extend the definition of a *signature* from binary languages to r-ary languages for $r > 2$ in the expected way:

Definition 6.4. The *signature* of a state description Θ as in (6.1) (or (6.2)) is defined to be the vector $\langle n_1, \ldots, n_p \rangle$, where n_j is the number of $\langle i_1, \ldots, i_r \rangle$ such that $1 \leq i_1 < \ldots < i_r \leq m$ and $\gamma_{h_{i_1, \ldots, i_r}} \in \Gamma_j$.

Thus, the signature records how many atoms from each equivalence class there are within $\Theta(b_1, \ldots, b_m)$. When $m < r$, the signature is not defined, but the notion of *extended signature* still makes sense, where the extended signature of Θ as in (6.2) is the vector

$$\langle n_1^1, \ldots, n_{p_1}^1 \,;\, \ldots \,;\, n_1^{r-1}, \ldots, n_{p_{r-1}}^{r-1} \,;\, n_1, \ldots, n_p \rangle\,,$$

and n_j^g is the number of $\langle i_1, \ldots, i_g \rangle$ such that $1 \leq i_1 < \ldots < i_g \leq m$ and $\gamma^g_{h_{i_1, \ldots, i_g}} \in \Gamma_j^g$. Note that the extended signature is derivable from the signature (when $m \geq r$) and that it is defined even when $m < r$.

Signature Exchangeability Principle, Sgx

The probability of a state description depends only on its signature.

Sgx for unary or binary languages is the same as Ex or BEx respectively. Sgx implies Ex but the converse implication does not hold in general. We gave two examples of probability functions satisfying Ex but not Sgx (BEx) for $r = 2$ in the previous chapter (see page 75).

6.3 Polyadic Independence

In the binary case, we defined C^2_ϕ as the set of pairs of constants brought together instantiating a relation in a sentence ϕ (Definition 5.4). In a similar vein, the following definition captures exactly which sets of g *constants* are brought together instantiating a relation within a sentence:

Definition 6.5. For a sentence $\phi(b_1, \dots, b_m) \in SL$ we define C^s_ϕ to be the set of all sets $\{b_{k_1}, \dots, b_{k_s}\}$ with s elements such that all of $b_{k_1}, \dots, b_{k_s}$ appear in some $\pm R_d(b_{i_1}, \dots, b_{i_{r_d}})$, $d \in \{1, \dots, q\}$ featuring in ϕ.

We refer to C^s_ϕ as the *set of s-sets of constants appearing in* ϕ. For example, consider a language containing one binary relation symbol R_1 and one ternary relation symbol R_2. For

$$\phi = R_1(a_7, a_2) \vee R_2(a_4, a_2, a_4)$$

we have $C^1_\phi = \{\{a_2\}, \{a_7\}, \{a_4\}\}$, $C^2_\phi = \{\{a_2, a_7\}, \{a_2, a_4\}\}$ and $C^k_\phi = \emptyset$ for $k \geq 3$. Note that $\bigcup_{s=1}^r C^s_\phi$ is closed under taking subsets.

As we had in the binary case, a modification of the Disjunctive Normal Form Theorem yields the following lemma:

Lemma 6.6. *Let $\phi(b_1, \dots, b_m) \in QFSL$. Then $\phi(b_1, \dots, b_m)$ is equivalent to a disjunction of partial state descriptions as in* (6.4)*, with $C^s = C^s_\phi$ for $s = 1, \dots, r$.*

We are now in a position to formulate a general version of the Independence Principle from page 65 based on atoms, as we did on page 77 for $r = 2$. In this generalised version we require that the following holds for any $g < r$: if two quantifier-free sentences have no $(g+1)$-sets of constants in common then they are conditionally independent given a g-ary trace for the constants that they share.

Strong Independence Principle, SIP

Let L be an r-ary language and let $0 \leq g < r$. Assume that $\phi, \psi \in QFSL$ are such that

$$C_{\phi}^{g+1} \cap C_{\psi}^{g+1} = \emptyset$$

and let $b_1, \ldots, b_t$ be the constants that ϕ and ψ have in common (if any). Let Δ be a g-ary trace for the constants $b_1, \ldots, b_t$ when $t > 0$, and $\Delta = \top$ (tautology) if ϕ and ψ have no constants in common. Then

$$w(\phi \wedge \psi \,|\, \Delta) = w(\phi \,|\, \Delta) \cdot w(\psi \,|\, \Delta). \tag{6.6}$$

The Basic SIP Functions

Recall that for $g \leq r$, N_g is the number of g-atoms and p_g is the number of equivalence classes of g-atoms under $\sim$.

Let $\vec{Y} = \langle y_1^1, \ldots, y_{p_1}^1 \,;\, y_1^2, \ldots, y_{p_2}^2 \,;\, \ldots \,;\, y_1^r, \ldots, y_{p_r}^r \rangle$ be a vector of real numbers such that

$$0 \leq y_j^g \leq 1, \qquad \sum_{j=1}^{p_1} y_j^1 = 1,$$

and such that for $1 < g \leq r$ the following holds: For any $(g-1)$-ary trace ψ for $x_1, \ldots, x_g$,

$$\sum_j s_j^g \, y_j^g = 1 \tag{6.7}$$

where the sum is taken over those $j \in \{1, \ldots, p_g\}$ for which Γ_j^g contains some γ_h^g with the $(g-1)$-ary trace ψ.

We use $\mathbb{D}_L$ to denote the set of vectors satisfying the above conditions. In a bid to keep our formulae simpler, we will write

$$z_h^g = y_{j(h)}^g$$

where $j(h)$ is that j for which $\gamma_h^g \in \Gamma_j^g$. Note that (6.7) is the same as requiring

$$\sum_{(\gamma_h^g \restriction g-1) = \psi} z_h^g \;=\; 1. \tag{6.8}$$

The vectors $\vec{Y} \in \mathbb{D}_L$ play a similar role in the polyadic to the role the vectors $\vec{x} \in \mathbb{D}_{2^q}$ from $w_{\vec{x}}$ play in the unary. For a given $\vec{Y}$, the corresponding function $w_{\vec{Y}}$ assigns a state description $\Theta(b_1, \ldots, b_m)$ the probability of obtaining it by the following process: First the γ_h^1 are chosen for $b_1, \ldots, b_m$, independently according to the probabilities z_h^1. Then the γ_h^2 are chosen for b_{i_1}, b_{i_2} with $i_1 < i_2$ from amongst the eligible ones, i.e. from amongst those γ_h^2 for which $(\gamma_h^2 \restriction 1)(x_1, x_2) \equiv \gamma_{h_{i_1}}^1(x_1) \wedge \gamma_{h_{i_2}}^1(x_2)$, independently and according to the probabilities z_h^2, and so on. Note that this works by virtue of (6.8), because when choosing γ_h^g for $b_{i_1}, \ldots, b_{i_g}$, $(\gamma_h^g \restriction g-1)$ is determined.

More formally, given $\vec{Y}$ as above, for a state description $\Theta(a_1, \ldots, a_m)$ such that

$$\Theta(a_1, \ldots, a_m) \equiv \bigwedge_{\substack{1 \leq s \leq r \\ 1 \leq i_1 < \ldots < i_s \leq m}} \gamma_{h_{i_1, \ldots, i_s}}^s(a_{i_1}, \ldots, a_{i_s}) \tag{6.9}$$

we define

$$w_{\vec{Y}}(\Theta(a_1, \ldots, a_m)) = \prod_{\substack{1 \leq s \leq r \\ 1 \leq i_1 < \ldots < i_s \leq m}} z_{h_{i_1, \ldots, i_s}}^s. \tag{6.10}$$

Note that, as in the binary case, if $\sigma \in \mathsf{S}_m$ and $\Psi(a_1, \ldots, a_m) = \Theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)})$ then $w_{\vec{Y}}(\Theta) = w_{\vec{Y}}(\Psi)$:

$$\Psi(a_1, \ldots, a_m) = \bigwedge_{\substack{1 \leq s \leq r \\ 1 \leq i_1 < \cdots < i_s \leq m}} \gamma_{h_{\sigma^{-1}(i_1), \ldots, \sigma^{-1}(i_s)}}^s(a_{i_1}, \ldots, a_{i_s})$$

and the multisets

$$\{\gamma_{h_{\sigma^{-1}(i_1), \ldots, \sigma^{-1}(i_s)}}^s : 1 \leq i_1 < \ldots i_s \leq m\}, \quad \{\gamma_{h_{i_1, \ldots, i_s}}^s : 1 \leq i_1 < \cdots < i_s \leq m\}$$

can only differ in which atoms from each equivalence class they contain, but not the number of atoms they contain from each equivalence class.

Proposition 6.7. *Let L be an r-ary language. The functions $w_{\vec{Y}}$ defined in* (6.10) *determine probability functions on SL that satisfy Ex.*

Proof. To show that $w_{\vec{Y}}$ determines a probability function note that (P1′) and (P2′) from page 13 clearly hold. For (P3′), we will prove that for any state description $\Theta(a_1, \ldots, a_m)$ we have

$$w_{\vec{Y}}(\Theta(a_1, \ldots, a_m)) = \sum_{\Theta^+(a_1, \ldots, a_m, a_{m+1}) \models \Theta(a_1, \ldots, a_m)} w_{\vec{Y}}(\Theta^+(a_1, \ldots, a_m, a_{m+1})).$$

Let $\Theta^+(a_1,\dots,a_m,a_{m+1})$ extend Θ. Then $w_{\vec{Y}}(\Theta^+(a_1,\dots,a_m,a_{m+1}))$ is the product

$$\left(\prod_{\substack{1\le s\le r\\ 1\le i_1<\dots<i_s\le m}} z^s_{h_{i_1,\dots,i_s}}\right)\left(\prod_{\substack{1\le s\le r\\ 1\le i_1<\dots<i_{s-1}\le m}} z^s_{h_{i_1,\dots,i_{s-1},(m+1)}}\right)$$

where the first product is as for Θ and $h_{i_1,\dots,i_{s-1},(m+1)}$ is that h for which

$$\gamma^s_h(a_{i_1},\dots,a_{i_{s-1}},a_{m+1}) = \Theta^+[a_{i_1},\dots,a_{i_{s-1}},a_{m+1}]\,.$$

That is, where Θ^+ is

$$\Theta(a_1,\dots,a_m) \quad\wedge \bigwedge_{\substack{1\le s\le r\\ 1\le i_1<\dots<i_{s-1}\le m}} \gamma^s_{h_{i_1,\dots,i_{s-1},(m+1)}}(a_{i_1},\dots,a_{i_{s-1}},a_{m+1}). \tag{6.11}$$

Consider some r-tuple $\langle i_1,\dots,i_{r-1},(m+1)\rangle$ with $1\le i_1<\dots<i_{r-1}\le m$. If some $\Theta^+\models\Theta$ satisfies

$$\Theta^+[a_{i_1},\dots,a_{i_{r-1}},a_{m+1}] = \gamma^r_h(a_{i_1},\dots,a_{i_{r-1}},a_{m+1})\,,$$

then any conjunction that differs from (6.11) only by having $\gamma^r_k(a_{i_1},\dots,a_{i_{r-1}},a_{m+1})$ in place of $\gamma^r_h(a_{i_1},\dots,a_{i_{r-1}},a_{m+1})$, where γ^r_h and γ^r_k have the same $(r-1)$-ary trace, is also a state description extending Θ. Since the z^r_k for all such k sum to 1 (from (6.8)), we can sum them out. Similarly, we can deal with the other r-tuples, then the $(r-1)$-tuples and so on, working our way down. So (P3$'$) holds too, and the $w_{\vec{Y}}$ define probability functions on SL.

The $w_{\vec{Y}}$ satisfy Ex by the remark preceding this proposition, since $w_{\vec{Y}}(\Theta(a_1,\dots,a_m)) = w_{\vec{Y}}(\Theta(a_{\sigma(1)},\dots,a_{\sigma(m)}))$ for state descriptions Θ and $\sigma\in\mathsf{S}_m$. $\square$

Proposition 6.8. *The $w_{\vec{Y}}$ satisfy Sgx.*

Proof. Similar reasoning to above (and as in the binary case) - summing the probabilities of all state descriptions for $a_1,\dots,a_M$ extending $\Theta(b_1,\dots,b_m)$ (where $b_1,\dots,b_m$ are amongst $a_1,\dots,a_M$) - gives us that (6.10) holds even when $a_1,\dots,a_m$ are replaced by any other distinct constants $b_1,\dots,b_m$. Therefore $w_{\vec{Y}}(\Theta)$ is independent of the constants instantiating Θ and hence it satisfies Sgx, since the right hand side of (6.10) depends only on the signature of Θ. $\square$

Next we would like to show that the $w_{\vec{Y}}$ satisfy SIP. For this purpose we will use the following lemma:

Lemma 6.9. *Let*

$$\Phi(b_1,\dots,b_m) = \bigwedge_{1\le s\le r} \bigwedge_{\substack{\{b_{i_1},\dots,b_{i_s}\}\in C^s\\ i_1<\ldots<i_s}} \gamma^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s}) \tag{6.12}$$

be a partial state description of an r-ary language. Then

$$w_{\vec{Y}}(\Phi(b_1,\dots,b_m)) = \prod_{1\le s\le r} \prod_{\substack{\{b_{i_1},\dots,b_{i_s}\}\in C^s\\ i_1<\ldots<i_s}} z^s_{h_{i_1,\dots,i_s}}.$$

Proof. To prove the lemma we sum the probabilities of state descriptions $\Theta(b_1,\dots,b_m)$ extending $\Phi(b_1,\dots,b_m)$, exactly as in the proof of Lemma 5.7. $\square$

Theorem 6.10. *Let L be an r-ary language. The probability functions $w_{\vec{Y}}$ on SL satisfy SIP and hence also IP.*

Furthermore, any probability function satisfying Ex and SIP is equal to $w_{\vec{Y}}$ for some $\vec{Y}\in\mathbb{D}_L$.

Proof. Using Lemma 6.9, we will first show that SIP holds for partial state descriptions Φ,Ψ and then, employing Lemma 6.6, in general. Let Φ be as in (6.12) and let

$$\Psi(b_1,\dots,b_t,b_{m+1},\dots,b_{m+n}) = \bigwedge_{1\le s\le r} \bigwedge_{\substack{\{b_{i_1},\dots,b_{i_s}\}\in D^s\\ i_1<\cdots<i_s}} \gamma^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s}) \tag{6.13}$$

with D^s a set of s-element subsets of $\{b_1,\dots,b_t,b_{m+1},\dots,b_{m+n}\}$. Suppose that $C^{g+1}\cap D^{g+1}=\emptyset$ and that $b_1,\dots,b_t$ are the constants Φ and Ψ have in common. Let $\Delta(b_1,\dots,b_t)=\bigwedge_{1\le s\le g}\bigwedge_{1\le i_1<\cdots<i_s\le t}\gamma^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s})$ be a g-ary trace for $b_1,\dots,b_t$. If Δ is inconsistent with Φ or Ψ then

$$w(\Phi\wedge\Psi\,|\,\Delta) = w(\Phi\,|\,\Delta)\cdot w(\Psi\,|\,\Delta) \tag{6.14}$$

holds because both sides are 0. We show that (6.14) holds also when $\Delta\wedge\Phi\wedge\Psi$ is consistent. Suppose first that $\Phi,\Psi\models\Delta$.

Using that $\Phi\wedge\Psi$ is a partial state description we have

$$w_{\vec{Y}}(\Phi\wedge\Psi\,|\,\Delta) = \frac{\prod_{1\le s\le r}\prod_{\substack{\{b_{i_1},\dots,b_{i_s}\}\in C^s\cup D^s\\ i_1<\cdots<i_s}} z^s_{h_{i_1,\dots,i_s}}}{\prod_{1\le s\le g}\prod_{1\le i_1<\cdots<i_s\le t} z^s_{h_{i_1,\dots,i_s}}}$$

$$= \prod_{1\leq s\leq g} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s\cup D^s \\ i_1<\cdots<i_s \\ t<i_s}} z^s_{h_{i_1,\ldots,i_s}} \prod_{g+1\leq s\leq r} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s\cup D^s \\ i_1<\cdots<i_s}} z^s_{h_{i_1,\ldots,i_s}}$$

and

$$w_{\vec{Y}}(\Phi \mid \Delta)\cdot w_{\vec{Y}}(\Psi \mid \Delta) = \frac{\prod_{1\leq s\leq r} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s \\ i_1<\ldots<i_s}} z^s_{h_{i_1,\ldots,i_s}}}{\prod_{1\leq s\leq g} \prod_{1\leq i_1<\cdots<i_s\leq t} z^s_{h_{i_1,\ldots,i_s}}} \cdot \frac{\prod_{1\leq s\leq r} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in D^s \\ i_1<\cdots<i_s}} z^s_{h_{i_1,\ldots,i_s}}}{\prod_{1\leq s\leq g} \prod_{1\leq i_1<\cdots<i_s\leq t} z^s_{h_{i_1,\ldots,i_s}}}$$

$$= \left(\prod_{1\leq s\leq g} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s \\ i_1<\cdots<i_s \\ t<i_s}} z^s_{h_{i_1,\ldots,i_s}} \prod_{g+1\leq s\leq r} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s \\ i_1<\cdots<i_s}} z^s_{h_{i_1,\ldots,i_s}} \right)$$

$$\cdot \left(\prod_{1\leq s\leq g} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in D^s \\ i_1<\cdots<i_s \\ t<i_s}} z^s_{h_{i_1,\ldots,i_s}} \prod_{g+1\leq s\leq r} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in D^s \\ i_1<\cdots<i_s}} z^s_{h_{i_1,\ldots,i_s}} \right)$$

$$= \prod_{1\leq s\leq g} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s\cup D^s \\ i_1<\cdots<i_s \\ t<i_s}} z^s_{h_{i_1,\ldots,i_s}} \prod_{g+1\leq s\leq r} \prod_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s\cup D^s \\ i_1<\cdots<i_s}} z^s_{h_{i_1,\ldots,i_s}},$$

so (6.14) holds for these Φ, Ψ. Now suppose it is not the case that $\Phi, \Psi \models \Delta$. Then for $1 \leq s \leq g$, there are $\gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1},\ldots,b_{i_s})$ which appear in Δ but do not appear in at least one of Φ, Ψ. We can see that (6.14) holds in this case too, since the additional factors in $w_{\vec{Y}}(\Phi\wedge\Psi\wedge\Delta)$, $w_{\vec{Y}}(\Phi\wedge\Delta)$ and $w_{\vec{Y}}(\Psi\wedge\Delta)$ are all divided out by $w_{\vec{Y}}(\Delta)$, so we end up with the same products as above. We conclude that SIP holds for partial state descriptions.

To see that it holds also for any $\phi, \psi \in QFSL$, using Lemma 6.6, let $\phi = \bigvee_u \Phi_u$ for partial state descriptions Φ_u as in (6.12) with $C^s = C^s_\Phi$, and $\psi = \bigvee_f \Psi_f$ for partial state descriptions Ψ_f as in (6.13) with $D^s = C^s_\Psi$. Then the required result follows as in the proof of Theorem 5.8 on page 85.

To prove the last part of the theorem, assume that w satisfies Ex and SIP. We define $\vec{Y}$ by

$$y^g_{j(h)} = z^g_h = w(\gamma^g_h(a_1,\ldots,a_g) \mid (\gamma^g_h \restriction g-1)(a_1,\ldots,a_g))$$

where $\gamma^g_h \in \Gamma^g_j$ and $(\gamma^g_h \restriction g-1)(a_1,\ldots,a_g)$ stands for a tautology when $g=1$. Note that by Ex it does not matter which γ^g_h from Γ^g_j we take, and that (6.7) must hold. Writing any state description in the form (6.9) and using Ex and SIP, we can show by induction[2] (adding the conjuncts for increasing numbers of constants one by one) that its probability is given by (6.10). □

Corollary 6.11. *Let L be an r-ary language and let μ be a normalised σ-additive measure on the Borel subsets of $\mathbb{D}_L$. For any $\theta \in SL$ define*

$$w(\theta) \;=\; \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\, d\mu(\vec{Y}). \tag{6.15}$$

Then the function w is a probability function on SL satisfying Sgx.

Proof. The function w defined on SL by (6.15) clearly satisfies (P1) and (P2) from page 12, and using Lebesgue's Dominated Convergence Theorem it also satisfies (P3). So w is a probability function. It satisfies Sgx since all the $w_{\vec{Y}}$ do. □

On the other hand, whether or not the converse to Corollary 6.11 holds, that is, whether any probability function satisfying Sgx can be expressed in the form (6.15), remains to be investigated. We know that such a representation theorem, if it exists, must use a different method to our proof of the binary Representation Theorem. We demonstrate why this is so on the case when L is a ternary language.

Recall that the proof of the binary Representation Theorem 5.9 relied on counting the number of state descriptions with a particular extended signature $\vec{u}\vec{t}$ (see (5.26)). This involved expressing the sums of the t_j for those j such that Γ_j contains an atom with a particular unary trace (that is the number of 'places' in the state description to be filled with binary atoms from these classes Γ_j) in terms of the u_k from $\vec{u}$, the unary part of this extended signature.

Let

$$\vec{m}\vec{n}\vec{l} = \langle m_1, \dots, m_{2^q}\,;\, n_1, \dots, n_{p_2}\,;\, l_1, \dots, l_p\rangle$$

denote an extended signature of some state description of the ternary language L. We will show that it is not possible to count the number of state descriptions with this extended signature in the way it is done in the binary and thus conclude that the method of the binary proof does not work in the general polyadic case. In particular, we will show that no expression for the l_j can be worked out in terms of the n_j; the

[2] As in the proof of Theorem 5.8.

number of binary atoms from each equivalence class (that is, $\vec{n}$) does not suffice to determine the number of 'places' which atoms with a given binary trace must fill.

To see this, notice first that for the extended signature $\vec{m}\vec{n}\vec{l}$ we still have, as in the binary:

$$\sum_{1\leq k\leq 2^q} m_k = m;$$

for $k \neq c$

$$\sum_j n_j = m_k m_c$$

where the sum is taken over those $j \in \{1,\ldots,p_2\}$ for which Γ^2_j contains some $\gamma^2_h(x_1,x_2)$ with unary trace $\gamma^1_k(x_1)\wedge\gamma^1_c(x_2)$, and

$$\sum_j n_j = \frac{m_k(m_k-1)}{2}$$

for those j such that Γ^2_j contains a $\gamma^2_h(x_1,x_2)$ with unary trace $\gamma^1_k(x_1)\wedge\gamma^1_k(x_2)$.

However, when we fix a binary trace ψ and sum the l_j over those $j \in \{1,\ldots,p\}$ such that Γ_j contains a $\gamma_h(x_1,x_2,x_3)$ with this binary trace ψ, different state descriptions of L may yield different results.

Let

$$\Theta(b_1,b_2,b_3,b_4) = \gamma_{h_{1,2,3}}(b_1,b_2,b_3)\wedge\gamma_{h_{1,2,4}}(b_1,b_2,b_4)\wedge\gamma_{h_{1,3,4}}(b_1,b_3,b_4)\wedge\gamma_{h_{2,3,4}}(b_2,b_3,b_4),$$
$$\Phi(b_1,b_2,b_3,b_4) = \gamma_{f_{1,2,3}}(b_1,b_2,b_3)\wedge\gamma_{f_{1,2,4}}(b_1,b_2,b_4)\wedge\gamma_{f_{1,3,4}}(b_1,b_3,b_4)\wedge\gamma_{f_{2,3,4}}(b_2,b_3,b_4)$$

be state descriptions of L such that

$$\begin{aligned}\Theta \models & \bigwedge_{i=1}^{4}\gamma^1_1(b_i)\wedge\gamma^2_1(b_1,b_2)\wedge\gamma^2_2(b_1,b_3)\wedge\gamma^2_2(b_1,b_4)\\ & \wedge\gamma^2_1(b_2,b_3)\wedge\gamma^2_2(b_2,b_4)\wedge\gamma^2_1(b_3,b_4),\end{aligned} \tag{6.16}$$

$$\begin{aligned}\Phi \models & \bigwedge_{i=1}^{4}\gamma^1_1(b_i)\wedge\gamma^2_1(b_1,b_2)\wedge\gamma^2_1(b_1,b_3)\wedge\gamma^2_2(b_1,b_4)\\ & \wedge\gamma^2_1(b_2,b_3)\wedge\gamma^2_2(b_2,b_4)\wedge\gamma^2_2(b_3,b_4)\end{aligned} \tag{6.17}$$

and where $\gamma^2_1(x_1,x_2)\not\sim\gamma^2_2(x_1,x_2)$ (so γ^2_1 and γ^2_2 are from different Γ^2_j).

Then Θ and Φ both have the same $\vec{m}$ and $\vec{n}$ parts of the extended signature, since each contains 4 of γ^1_1, 3 of γ^2_1 and 3 of γ^2_2 (i.e. Θ and Φ agree on all the m_k and n_j).

However, if we let ψ be the binary trace[3]

$$\gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_2^2(x_2,x_3)$$

and sum the l_j over those j for which Γ_j contains an atom with the binary trace ψ, we get different results for Θ and Φ. For Θ we have that

$$\sum_j l_j = 2$$

since (from (6.16)) the binary traces of the atoms appearing in Θ are as follows:

$$\begin{aligned}
\gamma_{h_{1,2,3}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_1^2(x_2,x_3),\\
\gamma_{h_{1,2,4}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_2^2(x_2,x_3),\\
\gamma_{h_{1,3,4}}(x_1,x_2,x_3) &\models \gamma_2^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_1^2(x_2,x_3),\\
\gamma_{h_{2,3,4}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_1^2(x_2,x_3),
\end{aligned}$$

and of these, only $\gamma_{h_{1,2,4}}$ and $\gamma_{h_{1,3,4}}$ are from a Γ_j that contains an atom with the binary trace ψ. In contrast, (using (6.17)) the binary traces of the atoms appearing in Φ are

$$\begin{aligned}
\gamma_{f_{1,2,3}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_1^2(x_1,x_3)\wedge\gamma_1^2(x_2,x_3)\\
\gamma_{f_{1,2,4}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_2^2(x_2,x_3)\\
\gamma_{f_{1,3,4}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_2^2(x_2,x_3)\\
\gamma_{f_{2,3,4}}(x_1,x_2,x_3) &\models \gamma_1^2(x_1,x_2)\wedge\gamma_2^2(x_1,x_3)\wedge\gamma_2^2(x_2,x_3)
\end{aligned}$$

so

$$\sum_j l_j = 3$$

for those j as above, since $\gamma_{f_{1,2,4}}$, $\gamma_{f_{1,3,4}}$ and $\gamma_{f_{2,3,4}}$ are all from a Γ_j that contains an atom with the binary trace ψ.

It follows that no expression using the n_j for these sums can be found, since the n_j are the same for Θ and Φ. Therefore, we cannot use the same proof method as in Theorem 5.9. Whether a different method may yield such a representation theorem remains, as already mentioned, a subject for future investigations.

[3]For clarity, we do not write the 1-atoms in the binary traces that follow; they are all $\gamma_1^1(x_i)$.

6.4 Polyadic Instantial Relevance

For a general r-ary language, instantial relevance based on atoms can be captured similarly to the binary case. To do this, we first generalise the concept of *Extra* to describe how much information a g-atom instantiated by $b_1, \ldots, b_g$ adds to a partial state description.

Let

$$\Delta(a_1, \ldots, a_m) \;=\; \bigwedge_{1 \leq s \leq r} \;\; \bigwedge_{\substack{\{a_{i_1}, \ldots, a_{i_s}\} \in C^s \\ i_1 < \ldots < i_s}} \gamma^s_{h_{i_1, \ldots, i_s}}(a_{i_1}, \ldots, a_{i_s}) \tag{6.18}$$

be a partial state description. Recall that $\bigcup_{s=1}^r C^s$ is assumed to be closed under taking subsets. Let $b_1, \ldots, b_g$ be distinct constants, some of which may be amongst $a_1, \ldots, a_m$. Assume that $\gamma^g_h(b_1, \ldots, b_g)$ is consistent with Δ.

Definition 6.12. The *Extra in* $\gamma^g_h(b_1, \ldots, b_g)$ *over* Δ is the set E of those subsets $\{t_1, \ldots, t_s\}$ of $\{1, \ldots, g\}$ such that $\{b_{t_1}, \ldots, b_{t_s}\}$ is not in $\bigcup_{s=1}^r C^s$.

Note that E is empty just if $\Delta(a_1, \ldots, a_m)$ implies $\gamma^g_h(b_1, \ldots, b_g)$, otherwise $\{1, \ldots, g\}$ is in E. E contains the singleton $\{i\}$ just when b_i is a new constant not featuring in Δ. E is the whole power set of $\{1, \ldots, g\}$ when all of $b_1, \ldots, b_g$ are new. The Extra is closed under supersets, and the additional information in $\gamma^g_h(b_1, \ldots, b_g)$ over Δ consists of all $\pm R_d(b_{i_1}, \ldots, b_{i_{r_d}})$ implied by $\gamma^g_h(b_1, \ldots, b_g)$ and such that $\{i_1, \ldots, i_{r_d}\} \in E$.

Polyadic Principle of Instantial Relevance, PPIR

Let $\Delta(a_1, \ldots, a_m)$ be a partial state description, $1 \leq g \leq r$, and let γ^g_h be a g-atom. Let $b_1, \ldots, b_g$, $b'_1, \ldots, b'_g$ be such that

$$\Delta \wedge \gamma^g_h(b_1, \ldots, b_g) \wedge \gamma^g_h(b'_1, \ldots, b'_g)$$

is consistent. Assume that the Extras in $\gamma^g_h(b_1, \ldots, b_g)$ over $\Delta \wedge \gamma^g_h(b'_1, \ldots, b'_g)$, in $\gamma^g_h(b_1, \ldots, b_g)$ over Δ and in $\gamma^g_h(b'_1, \ldots, b'_g)$ over Δ are all the same. Then

$$w(\gamma^g_h(b_1, \ldots, b_g) \,|\, \Delta) \leq w(\gamma^g_h(b_1, \ldots, b_g) \,|\, \gamma^g_h(b'_1, \ldots, b'_g) \wedge \Delta)\,. \tag{6.19}$$

Theorem 6.13. *Any convex combination (or integral) of the functions $w_{\vec{Y}}$ satisfies PPIR.*

Proof. Let Δ, γ_h^g and $b_1, \ldots, b_g$, $b_1', \ldots, b_g'$ be as in the statement of PPIR. Assume Δ is as in (6.18). Let E be the Extra in $\gamma_h^g(b_1, \ldots, b_g)$ over Δ. We have

$$w_{\vec{Y}}(\Delta) = \prod_{1 \le s \le r} \prod_{\substack{\{a_{i_1}, \ldots, a_{i_s}\} \in C^s \\ i_1 < \ldots < i_s}} z^s_{h_{i_1, \ldots, i_s}},$$

$$w_{\vec{Y}}(\Delta \wedge \gamma_h^g(b_1, \ldots, b_g)) = w_{\vec{Y}}(\Delta \wedge \gamma_h^g(b_1', \ldots, b_g')) = w_{\vec{Y}}(\Delta) \cdot \prod_{\{t_1, \ldots, t_s\} \in E} z^s_{k_{t_1, \ldots, t_s}},$$

$$w_{\vec{Y}}(\Delta \wedge \gamma_h^g(b_1, \ldots, b_g) \wedge \gamma_h^g(b_1', \ldots, b_g')) = w_{\vec{Y}}(\Delta) \cdot \left(\prod_{\{t_1, \ldots, t_s\} \in E} z^s_{k_{t_1, \ldots, t_s}} \right)^2.$$

It follows that for $w = w_{\vec{Y}}$, (6.19) holds with equality.

The proof for w defined by (6.15), and hence also for any convex combination of the $w_{\vec{Y}}$, follows from the above equations exactly as in the binary case (Theorem 5.13). $\square$

By the same method we also obtain that under the same assumptions as those in PPIR except that merely the Extras in $\gamma_h^g(b_1, \ldots, b_g)$ over $\Delta \wedge \gamma_h^g(b_1', \ldots, b_g')$ and in $\gamma_h^g(b_1', \ldots, b_g')$ over Δ are required to be the same, we obtain that any convex combination (or integral) w of the functions $w_{\vec{Y}}$ satisfies

$$w(\gamma_h^g(b_1', \ldots, b_g') \mid \Delta) \le w(\gamma_h^g(b_1, \ldots, b_g) \mid \gamma_h^g(b_1', \ldots, b_g') \wedge \Delta).$$

6.5 Future Directions

The results presented in this chapter open the door to many new questions arising from these ideas. For example, we introduced the concept of g-atoms for $g \le r$ (where r is the arity of the language) and used it to define partial state descriptions. Considering a fixed r-ary language, let g-atoms again be the state formulae for g variables, but this time for any positive natural number g. We may define the *g-signature* of a state description for m individuals (where $g \le m$) analogously to (r-)signatures. We end this chapter with some observations regarding these g-signatures, a direction to be further researched.

We can see that the g-signature of a state description determines its s-signature for $s < g$. Hence, for such s, g, a probability function which gives state descriptions with the same s-signature the same probability, must also give the same probability to state descriptions with the same g-signature.

Conversely, however, it is not the case that the s-signature of a state description determines its g-signature for $s < g$, not even when $r \leq s < g$. One example, for $r = 2$, $s = 2$ and $g = 3$, is provided by the state descriptions on page 74. Here we give another example, for $r = 2$, $s = 3$ and $g = 4$:

Example. Let L contain one binary relation symbol. Then the 6 state formulae (3-atoms) represented by

$$\begin{array}{ccc} 1&1&1\\ 0&1&1\\ 0&0&1 \end{array}\qquad \begin{array}{ccc} 1&0&0\\ 1&1&0\\ 1&1&1 \end{array}\qquad \begin{array}{ccc} 1&1&0\\ 0&1&0\\ 1&1&1 \end{array}\qquad \begin{array}{ccc} 1&0&1\\ 1&1&1\\ 0&0&1 \end{array}\qquad \begin{array}{ccc} 1&0&0\\ 1&1&1\\ 1&0&1 \end{array}\qquad \begin{array}{ccc} 1&1&1\\ 0&1&0\\ 0&1&1 \end{array}$$

are equivalent in the sense that they can be obtained from each other via a permutation of the variables. Furthermore, the following two are also equivalent:

$$\begin{array}{ccc} 1&1&0\\ 0&1&1\\ 1&0&1 \end{array}\qquad \begin{array}{ccc} 1&0&1\\ 1&1&0\\ 0&1&1 \end{array}.$$

However, the state descriptions Θ and Φ represented respectively by the matrices

$$\begin{array}{cccccc} 1&1&0&1&1&1\\ 0&1&1&1&1&1\\ 1&0&1&1&1&1\\ 0&0&0&1&1&0\\ 0&0&0&0&1&1\\ 0&0&0&1&0&1 \end{array}\qquad \begin{array}{cccccc} 1&1&1&0&1&1\\ 0&1&1&1&1&1\\ 0&0&1&1&1&1\\ 1&0&0&1&1&1\\ 0&0&0&0&1&1\\ 0&0&0&0&0&1 \end{array}$$

feature only the above 3-atoms and they have the same 3-signature but not the same 4-signature. The 3-signature of both contains eighteen 3-atoms of the first kind and two 3-atoms of the second kind. On the other hand, the 4-atoms appearing in Θ and Φ, where those equivalent to each other appear on the same line, are

1.

$$\begin{matrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{matrix}$$

2.

$$\begin{matrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{matrix} \qquad \begin{matrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{matrix} \qquad \begin{matrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{matrix} \qquad \begin{matrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{matrix}$$

3.

$$\begin{matrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{matrix}$$

4.

$$\begin{matrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{matrix}.$$

The 4-signature of Θ consists of three 4-atoms of the first kind, nine of the second kind, and three of the third kind. However, the 4-signature of Φ contains four 4-atoms of the first kind, ten of the second kind, and one of the fourth kind. Therefore, the 4-signatures of Θ and Φ are not equal.

Chapter 7

Element-based Signature Exchangeability

7.1 Introduction

In the previous two chapters we presented an approach to generalising unary results involving exchangeability which was based on the concept of polyadic atoms. In this chapter, we propose an alternative approach based instead on the notion of *elements*. Elements are 'smaller' building blocks than atoms in the sense that the polyadic atoms can themselves be built out of elements. Rather than specifying how a set of constants behaves within a state description, an element describes the behaviour of an *ordered tuple* of constants within a state description. In the unary context, these two notions clearly coincide: a state formula on one variable is an element as well as an atom. On the other hand, when the language is non-unary, these notions differ considerably.

We shall see that the element-based approach generalises more readily to any polyadic language, and that the mathematics of the theory is in some ways more similar to the unary theory than in the atom-based approach. In addition, state descriptions expressed in terms of elements contain no redundancy, like state descriptions in the unary context. The drawback however, is that this approach is completely 'disjointed' - the element instantiated by an ordered tuple of constants is entirely independent from the element instantiated by the same constants permuted. This overlooks the interconnections that exist between constants in the polyadic (and do not exist in the unary, since a unary predicate acts on a single constant and cannot bring different constants together). Furthermore, we will see that the element-based approach takes us further away from polyadic Ex, a desirable property to have.

In the coming sections, we will define the elements of a polyadic language, the notion of an element-based signature, and a principle of invariance under such signatures. We will introduce functions that satisfy this principle, and provide a representation for *all* functions satisfying it. Finally, we will provide a principle of instantial relevance for this context.

7.2 An Element-based Approach for Polyadic Languages

Throughout this chapter (unless stated otherwise) let L be an r-ary language containing q relation symbols $R_1, R_2, \dots, R_q$ of arities $r_1, r_2, \dots, r_q$ respectively. Define $q_s = |\{d \,|\, r_d = s\}|$ for $d \in \{1, \dots, q\}$. That is, q_s denotes how many of $R_1, \dots, R_q$ have arity s.

Definition 7.1. For each $s \in \{1, \dots, r\}$ such that $q_s \neq 0$, the *s-ary elements* of L are the formulae of the form[1]

$$\bigwedge_{\substack{d \in \{1,\dots,q\} \\ r_d = s}} \pm R_d(x_1, \dots, x_s).$$

They are denoted by $\delta^s_1(x_1, \dots, x_s), \dots, \delta^s_{2^{q_s}}(x_1, \dots, x_s)$.

We order the s-ary elements lexicographically, so $\delta^s_1(x_1, \dots, x_s) = \bigwedge_{\substack{d \in \{1,\dots,q\} \\ r_d = s}} R_d(x_1, \dots, x_s)$ and $\delta^s_{2^{q_s}}(x_1, \dots, x_s) = \bigwedge_{\substack{d \in \{1,\dots,q\} \\ r_d = s}} \neg R_d(x_1, \dots, x_s)$.

A state description $\Theta(b_1, \dots, b_m)$ of L can be written as a conjunction of elements:

$$\bigwedge_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \; \bigwedge_{\langle i_1,\dots,i_s\rangle \in \{1,\dots,m\}^s} \delta^s_{h_{i_1,\dots,i_s}}(b_{i_1}, \dots, b_{i_s}), \tag{7.1}$$

[1] As usual, $x_1, \dots, x_s$ denote distinct variables.

with $h_{i_1,\dots,i_s} \in \{1,\dots,2^{q_s}\}$. Notice that every such conjunction defines a state description of L since it is necessarily consistent and it completely describes $b_1,\dots,b_m$ in relation to each other (and no other constants). Furthermore, the conjunction (7.1) contains *no redundancy*; we must specify $\delta^s_{h_{i_1,\dots,i_s}}$ for every s-tuple $\langle i_1,\dots i_s\rangle \in \{1,\dots,m\}^s$ to produce a state description.

Using the s-ary elements, we define the concept of *e-partial state descriptions*:

Definition 7.2. An *e-partial state description* of L is any consistent conjunction of elements of the form

$$\Delta(b_1,\dots,b_m) = \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in A^s} \delta^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s}) \tag{7.2}$$

where $A^s \subseteq \{1,\dots,m\}^s$.

e-partial state formulae are defined analogously, and we follow the convention that only the individuals which appear after some $\delta^s_{h_{i_1,\dots,i_s}}$ in (7.2) are listed in brackets after Δ and that those listed are distinct. We remark that this definition is not equivalent to the definition of a partial state description in the atom-based approach, see page 106.

7.3 e-Signatures

Definition 7.3. We define the *e-signature* of a state description

$$\Theta(b_1,\dots,b_m) = \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m\}^s} \delta^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s}),$$

to be the vector

$$\vec{M} = \langle m^s_1,\dots,m^s_{2^{q_s}} \;:\; s\in\{1,\dots,r\} \text{ and } q_s\neq 0\rangle,$$

where m^s_k is the number of times δ^s_k appears amongst the $\delta^s_{h_{i_1,\dots,i_s}}$.

For each $s\in\{1,\dots,r\}$ with $q_s\neq 0$ we have[2]

$$\sum_{k=1}^{2^{q_s}} m^s_k = m^s \tag{7.3}$$

[2]Note that where m has no subscript, m^s denotes *m to the power s*.

since there are m^s many s-tuples in $\{1,\dots,m\}^s$ and a state description specifies an s-ary element for each one. Conversely, every $\vec{M}$ with $m^s_k \in \mathbb{N}$ that satisfies (7.3) is an e-signature for some state description $\Theta(b_1,\dots,b_m)$ so we say it is an *e-signature on* m.

Notice that the definition of an e-signature may also be applied to e-partial state descriptions, however the sum in (7.3) holds only for (complete) state descriptions.

We now wish to define the concept of invariance under e-signatures. If the polyadic case worked like the unary we would have that this is equivalent to Ex as stated on page 64. However, we will shortly see that this in not the case and so we define the

e-Signature Exchangeability Principle, e-Sgx

The probability of a state description depends only on its e-signature.

e-Sgx implies Ex, since a state description's e-signature is invariant under permutations of constants. To see this, let $\Theta(b_1,\dots,b_m)$ be a state description as in (7.1) and suppose its e-signature is $\vec{M}$. Consider

$$\Theta(b_{\tau(1)},\dots,b_{\tau(m)}) = \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m\}^s} \delta^s_{h_{\tau^{-1}(i_1),\dots,\tau^{-1}(i_s)}}(b_{i_1},\dots,b_{i_s}),$$

the state description obtained from Θ by permuting $b_1,\dots,b_m$ according to $\tau \in \mathsf{S}_m$. Then the e-signature of $\Theta(b_{\tau(1)},\dots,b_{\tau(m)})$ is also $\vec{M}$, since $\Theta(b_{\tau(1)},\dots,b_{\tau(m)}) \models \delta^s_h(b_{i_1},\dots,b_{i_s})$ just if $\Theta(b_1,\dots,b_m) \models \delta^s_h(b_{\tau^{-1}(i_1)},\dots,b_{\tau^{-1}(i_s)})$, and so each $h \in \{1,\dots,2^{q_s}\}$ appears the same number of times amongst the $h_{i_1,\dots,i_s}$ and the $h_{\tau^{-1}(i_1),\dots,\tau^{-1}(i_s)}$.

On the other hand, two state descriptions with the same e-signature may not be obtainable from each other by permuting constants and can therefore get different probabilities from functions satisfying Ex. The two state descriptions from page 74 used to show this for BEx work here too. The e-signature of both is $\langle 10, 6\rangle$, but they get different probabilities from $u^{\bar{p},L}$ with $\bar{p} = \langle 0, \frac{1}{2}, \frac{1}{2}, 0, 0, \dots\rangle$.

The following result brings together the principles of Ex, Sgx[3] and e-Sgx.

Proposition 7.4. *Let L be an r-ary language, $r > 1$. Then e-Sgx $\Longrightarrow$ Sgx $\Longrightarrow$ Ex, however Ex $\not\Longrightarrow$ Sgx $\not\Longrightarrow$ e-Sgx.*

[3]stated on page 107.

Proof. We already have that Sgx implies Ex but not the converse (see page 74). To see that e-Sgx implies Sgx, we will show that if two state descriptions have the same signature then they also have the same e-signature, and hence any function that gives the same probability to state descriptions with the same e-signature would also satisfy Sgx. Consider an r-ary language L containing q relation symbols as above, and let $\Theta(b_1, \dots, b_m)$ be the state description

$$\bigwedge_{1 \leq i_1 < \cdots < i_r \leq m} \gamma_{h_{i_1,\dots,i_r}}(b_{i_1}, \dots, b_{i_r}). \tag{7.4}$$

Notice that for each $\gamma_{h_{i_1,\dots,i_r}}$ in (7.4) we have

$$\gamma_{h_{i_1,\dots,i_r}}(b_{i_1}, \dots, b_{i_r}) = \bigwedge_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \ \bigwedge_{\langle j_1,\dots,j_s \rangle \in \{i_1,\dots,i_r\}^s} \delta^s_{h_{j_1,\dots,j_s}}(b_{j_1}, \dots, b_{j_s})$$

since the γ_h are themselves state descriptions, so *all* the elements instantiated by constants from $b_{i_1}, \dots, b_{i_r}$ are determined by $\gamma_{h_{i_1,\dots,i_r}}$. Considering *every* $\gamma_{h_{i_1,\dots,i_r}}$ in (7.4) gives every s-ary element of the state description[4], therefore the e-signature of a state description is determined by its signature.

We now show the converse implication does not hold; state descriptions with the same e-signature may have different signatures and consequently obtain different probabilities from functions satisfying Sgx. Consider a language containing a single binary relation symbol and let Θ, Φ be the state descriptions represented respectively by the following matrices:

$$\begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix} \qquad\qquad \begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix}\ .$$

Let $w_{\vec{Y}}$ be a probability function as in (5.18) with $\vec{Y}$ defined by $x_1 = \frac{1}{4}$, $x_2 = \frac{3}{4}$; $y_1 = 1$, $y_7 = 1$, $y_{10} = 1$ and $y_j = 0$ otherwise, where[5]

$$\Gamma_1 = \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}, \quad \Gamma_7 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}, \quad \Gamma_{10} = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}.$$

Then Θ and Φ both have the e-signature $\langle 1, 3 \rangle$, however $w_{\vec{Y}}(\Theta) = \frac{3}{16}$, $w_{\vec{Y}}(\Phi) = 0$, and $w_{\vec{Y}}$ satisfies Sgx. $\square$

[4] In fact, the element satisfied by $\langle b_{j_1}, \dots, b_{j_s} \rangle$ will be implied by every $\gamma_{h_{i_1,\dots,i_r}}$ where $\{b_{j_1}, \dots, b_{j_s}\} \subseteq \{b_{i_1}, \dots, b_{i_r}\}$ so when counting the numbers of elements appearing in a state description using its signature, this repetition must be taken into account.

[5] The Γ_j were defined on page 72.

We remark that when the language is unary, the three principles e-Sgx, Sgx and Ex are all equivalent, since the signature and e-signature of a state description are equal.

7.4 The Probability Functions $W_{\vec{X}}$

The probability functions satisfying e-Sgx are closely related to the unary $w_{\vec{x}}$ and to the functions $w_{\vec{Y}}$ from the previous chapter. They are defined as follows.

Let $\mathbb{H}_L$ be the set of all

$$\vec{X} = \langle x_1^s, \ldots, x_{2^{q_s}}^s \;:\; s \in \{1, \ldots, r\} \text{ and } q_s \neq 0 \rangle$$

such that $x_k^s \in [0, 1]$ and for each s such that $q_s \neq 0$

$$\sum_{k=1}^{2^{q_s}} x_k^s = 1.$$

For a state description

$$\Theta(a_1, \ldots, a_m) = \bigwedge_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \; \bigwedge_{\langle i_1,\ldots,i_s \rangle \in \{1,\ldots,m\}^s} \delta^s_{h_{i_1,\ldots,i_s}}(a_{i_1}, \ldots, a_{i_s})$$

we define

$$W_{\vec{X}}(\Theta(a_1, \ldots, a_m)) = \prod_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \; \prod_{\langle i_1,\ldots,i_s \rangle \in \{1,\ldots,m\}^s} x^s_{h_{i_1,\ldots,i_s}} \tag{7.5}$$

where $h_{i_1,\ldots,i_s}$ is that $k \in \{1, \ldots, 2^{q_s}\}$ for which $\Theta \models \delta_k^s(a_{i_1}, \ldots, a_{i_s})$. We define the empty product to be 1.

Proposition 7.5. *The functions $W_{\vec{X}}$ defined in* (7.5) *determine probability functions on SL that satisfy Ex.*

Proof. We check that the $W_{\vec{X}}$ satisfy conditions (P1′)-(P3′) from page 13. (P1′) clearly holds. To see that (P2′) holds, let

$$\Theta_i(a_1) = \bigwedge_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \delta^s_{k_s}(a_1, a_1, \ldots, a_1)$$

be a state description on one constant, so

$$W_{\vec{X}}(\Theta_i(a_1)) = \prod_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} x^s_{k_s}$$

for $k_s \in \{1, \dots, 2^{q_s}\}$. There are $2^{q_1+\dots+q_r} = 2^q$ possibilities for Θ_i, hence

$$W_{\vec{X}}(\top) = W_{\vec{X}}\left(\bigvee_{i=1}^{2^q} \Theta_i(a_1)\right) = \sum_{i=1}^{2^q} W_{\vec{X}}(\Theta_i(a_1)) = \sum_{\vec{k}} \prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} x^s_{k_s} = \prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \sum_{k=1}^{2^{q_s}} x^s_k = 1$$

where $\vec{k} = \langle k_s : s \in \{1, \dots, r\},\ q_s \neq 0\rangle$ and each k_s in $\vec{k}$ is from $\{1, \dots, 2^{q_s}\}$.

To see that (P3′) holds, let $\Phi(a_1, \dots, a_{m+1}) \models \Theta(a_1, \dots, a_m)$. Then

$$\begin{aligned}\Phi(a_1, \dots, a_{m+1}) &= \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m+1\}^s} \delta^s_{h_{i_1,\dots,i_s}}(a_{i_1}, \dots, a_{i_s}) \\ &= \Theta(a_1, \dots, a_m) \wedge \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m+1\}^s\setminus\{1,\dots,m\}^s} \delta^s_{h_{i_1,\dots,i_s}}(a_{i_1}, \dots, a_{i_s})\end{aligned}$$

where $\{1, \dots, m+1\}^s \setminus \{1, \dots, m\}^s$ denotes the set of s-tuples from $\{1, \dots, m+1\}^s$ which are not also in $\{1, \dots, m\}^s$. Therefore for Θ, Φ as above

$$\begin{aligned}\sum_{\Phi\models\Theta} W_{\vec{X}}(\Phi) &= W_{\vec{X}}(\Theta) \cdot \sum_{\Phi\models\Theta} \left(\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m+1\}^s\setminus\{1,\dots,m\}^s} x^s_{h_{i_1,\dots,i_s}} \right) \\ &= W_{\vec{X}}(\Theta) \cdot \prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m+1\}^s\setminus\{1,\dots,m\}^s} \sum_{k=1}^{2^{q_s}} x^s_k \\ &= W_{\vec{X}}(\Theta)\end{aligned}$$

by swapping the order of the sum and products, and since for any s-tuple from $\{1, \dots, m+1\}^s \setminus \{1, \dots, m\}^s$ we may pick any δ^s_k and the resulting state description will be consistent with Θ. So (P3′) holds too and hence $W_{\vec{X}}$ extends to a probability function on SL.

To see that the $W_{\vec{X}}$ satisfy Ex, let $\tau \in \mathsf{S}_m$ and consider

$$\Theta(a_{\tau(1)}, \dots, a_{\tau(m)}) = \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m\}^s} \delta^s_{h_{i_1,\dots,i_s}}(a_{\tau(i_1)}, \dots, a_{\tau(i_s)}).$$

Then

$$W_{\vec{X}}(\Theta(a_{\tau(1)}, \dots, a_{\tau(m)})) = \prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,m\}^s} x^s_{h_{i_1,\dots,i_s}} = W_{\vec{X}}(\Theta(a_1, \dots, a_m)),$$

so we conclude that Ex holds, reasoning as we did in Proposition 5.5. $\square$

Proposition 7.6. *The $W_{\vec{X}}$ satisfy e-Sgx.*

Proof. Similarly to the proof of Proposition 5.6 and to the argument used in Proposition 7.5 above, we can replace $a_1, \dots, a_m$ in (7.5) by any distinct constants $b_1, \dots, b_m$ as follows. Let $a_1, \dots, a_M$ be such that all of $b_1, \dots, b_m$ are amongst them, and let $\Phi(a_1, \dots, a_M) \models \Theta(b_1, \dots, b_m)$. Then for such Φ, Θ

$$\sum_{\Phi(a_1,\dots,a_M) \models \Theta(b_1,\dots,b_m)} W_{\vec{X}}(\Phi(a_1, \dots, a_M)) = W_{\vec{X}}(\Theta(b_1, \dots, b_m))$$

since we have free choice over all the δ^s_k which do not appear in Θ, and for each s all such possibilities sum up to 1. We conclude that $W_{\vec{X}}(\Theta)$ is independent of the constants instantiating Θ, and so the right hand side of (7.5) depends only on the e-signature of Θ. □

The $w_{\vec{Y}}$ are more general than the $W_{\vec{X}}$, since every $W_{\vec{X}}$ is also a $w_{\vec{Y}}$ but not the converse. To show this is the case, we will need the following lemma:

Lemma 7.7. *Let L be an r-ary language and let*

$$\Delta(b_1, \dots, b_m) = \bigwedge_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \bigwedge_{\langle i_1,\dots,i_s \rangle \in A^s} \delta^s_{h_{i_1,\dots,i_s}}(b_{i_1}, \dots, b_{i_s}),$$

be an e-partial state description of L. Then

$$W_{\vec{X}}(\Delta(b_1, \dots, b_m)) = \prod_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \prod_{\langle i_1,\dots,i_s \rangle \in A^s} x^s_{h_{i_1,\dots,i_s}}.$$

Proof. Let $\Theta(b_1, \dots, b_m)$ be a state description extending $\Delta(b_1, \dots, b_m)$.

$$\begin{aligned}
W_{\vec{X}}(\Delta(b_1, \dots, b_m)) &= \sum_{\Theta \models \Delta} W_{\vec{X}}(\Theta(b_1, \dots, b_m)) \\
&= \sum_{\Theta \models \Delta} \left(\prod_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \left(\prod_{\langle i_1,\dots,i_s \rangle \in A^s} x^s_{h_{i_1,\dots,i_s}} \prod_{\langle i_1,\dots,i_s \rangle \in \{1,\dots,m\}^s \setminus A^s} x^s_{h_{i_1,\dots,i_s}} \right) \right) \\
&= \prod_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \prod_{\langle i_1,\dots,i_s \rangle \in A^s} x^s_{h_{i_1,\dots,i_s}} \cdot \sum_{\Theta \models \Delta} \left(\prod_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \prod_{\langle i_1,\dots,i_s \rangle \in \{1,\dots,m\}^s \setminus A^s} x^s_{h_{i_1,\dots,i_s}} \right) \\
&= \prod_{\substack{s \in \{1,\dots,r\} \\ q_s \neq 0}} \prod_{\langle i_1,\dots,i_s \rangle \in A^s} x^s_{h_{i_1,\dots,i_s}},
\end{aligned}$$

where $\{1,\dots,m\}^s \setminus A^s$ is the set of s-tuples from $\{1,\dots,m\}^s$ that are not in A^s. The final equality follows by swapping the order of the sum and products and since for the s-ary elements that do not appear in Δ, we may choose any δ^s_k, $k \in \{1,\dots,2^{q_s}\}$ and the resulting state description will extend Δ, and these choices sum up to 1. $\square$

We will also need the following observations. Firstly, recall our definition of the g-ary trace of a state description from page 105. We wish to write the g-ary trace of a state description Θ as in (7.1) using elements. For this purpose, let $\vec{j}$ denote the s-tuple $\langle j_1,\dots,j_s\rangle$ and define the function $f : \{1,\dots,m\}^s \to \{1,\dots,s\}$, where $f(\vec{j}) = n$ is the number of distinct constants amongst $j_1,\dots,j_s$. Then

$$(\Theta \restriction g)(b_1,\dots,b_m) = \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\substack{\langle j_1,\dots,j_s\rangle\in\{1,\dots,m\}^s\\ f(\vec{j})\leq g}} \delta^s_{h_{j_1,\dots,j_s}}(b_{j_1},\dots,b_{j_s}).$$

Notice that the g-ary trace of a state description is an e-partial state description. In what follows we will take the 0-ary trace of a state description to be a tautology.

Secondly, notice that if γ^g_h and γ^g_l are g-atoms such that $\gamma^g_h \sim \gamma^g_l$,[6] then

$$\begin{aligned} &\gamma^g_h(x_1,\dots,x_g) \equiv \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,g\}^s} \delta^s_{h_{i_1,\dots,i_s}}(x_{i_1},\dots,x_{i_s}) \\ \iff &\gamma^g_l(x_{\sigma(1)},\dots,x_{\sigma(g)}) \equiv \bigwedge_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \bigwedge_{\langle i_1,\dots,i_s\rangle\in\{1,\dots,g\}^s} \delta^s_{h_{i_1,\dots,i_s}}(x_{\sigma(i_1)},\dots,x_{\sigma(i_s)}) \end{aligned} \tag{7.6}$$

for some $\sigma \in \mathsf{S}_g$, so all the g-atoms in an equivalence class Γ^g_j logically imply the same s-ary elements.

Theorem 7.8. *Every $W_{\vec{X}}$ with $\vec{X} \in \mathbb{H}_L$ is one of the functions $w_{\vec{Y}}$ for some $\vec{Y} \in \mathbb{D}_L$, but not the converse.*

Proof. Let $\vec{X} = \langle x^s_1,\dots,x^s_{2^{q_s}} : s \in \{1,\dots,r\},\ q_s \neq 0\rangle \in \mathbb{H}_L$. By Lemma 7.7 and the first of the above observations we have (for $1 \leq g \leq r$)

$$\begin{aligned} W_{\vec{X}}(\gamma^g_h(a_1,\dots,a_g) \mid (\gamma^g_h \restriction g-1)(a_1,\dots,a_g)) &= \frac{\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{\langle j_1,\dots,j_s\rangle\in\{1,\dots,g\}^s} x^s_{h_{j_1,\dots,j_s}}}{\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{\substack{\langle j_1,\dots,j_s\rangle\in\{1,\dots,g\}^s\\ f(\vec{j})\leq g-1}} x^s_{h_{j_1,\dots,j_s}}} \\ &= \prod_{\substack{s\in\{g,\dots,r\}\\ q_s\neq 0}} \prod_{\substack{\langle j_1,\dots,j_s\rangle\in\{1,\dots,g\}^s\\ f(\vec{j})=g}} x^s_{h_{j_1,\dots,j_s}} \end{aligned}$$

[6] for the equivalence defined on page 107.

since we are left only with elements that contain g different constants, so each must instantiate all of $a_1, \ldots, a_g$ and hence must also be at least g-ary. Notice that since the $W_{\vec{X}}$ satisfy Ex, the above holds also when $a_1, \ldots, a_g$ are replaced with any other distinct constants.

Define $\vec{Y}$ by

$$y^g_{j(h)} = z^g_h = \prod_{\substack{s \in \{g,\ldots,r\} \\ q_s \neq 0}} \prod_{\substack{\langle j_1,\ldots,j_s\rangle \in \{1,\ldots,g\}^s \\ f(\vec{j})=g}} x^s_{h_{j_1,\ldots,j_s}} \tag{7.7}$$

where the $x^s_{h_{j_1,\ldots,j_s}}$ from $\vec{X}$ are such that $\gamma^g_h(x_1, \ldots, x_g) \models \delta^s_{h_{j_1,\ldots,j_s}}(x_{j_1}, \ldots, x_{j_s})$. Note that by the second observation, when $\gamma^g_h \sim \gamma^g_l$, we will get $z^g_h = z^g_l$. Then $\vec{Y} \in \mathbb{D}_L$ since (6.8) is satisfied and furthermore, if $\Theta(b_1, \ldots, b_m)$ is the state description

$$\bigwedge_{\substack{1 \leq g \leq r \\ 1 \leq i_1 < \cdots < i_g \leq m}} \gamma^g_{h_{i_1,\ldots,i_g}}(b_{i_1}, \ldots, b_{i_g}) \equiv \bigwedge_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \bigwedge_{\langle j_1,\ldots,j_s\rangle \in \{1,\ldots,m\}^s} \delta^s_{h_{j_1,\ldots,j_s}}(b_{j_1}, \ldots, b_{j_s})$$

then

$$\begin{aligned}
w_{\vec{Y}}(\Theta) &= \prod_{\substack{1 \leq g \leq r \\ 1 \leq i_1 < \cdots < i_g \leq m}} z^g_{h_{i_1,\ldots,i_g}} \\
&= \prod_{\substack{1 \leq g \leq r \\ 1 \leq i_1 < \cdots < i_g \leq m}} \left(\prod_{\substack{s \in \{g,\ldots,r\} \\ q_s \neq 0}} \prod_{\substack{\langle j_1,\ldots,j_s\rangle \in \{i_1,\ldots,i_g\}^s \\ f(\vec{j})=g}} x^s_{h_{j_1,\ldots,j_s}} \right) \\
&= \prod_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \prod_{\langle j_1,\ldots,j_s\rangle \in \{1,\ldots,m\}^s} x^s_{h_{j_1,\ldots,j_s}} \\
&= W_{\vec{X}}(\Theta)
\end{aligned}$$

using (6.10), (7.7) and (7.5). So for $\vec{Y}$ defined as above and every state description Θ, $W_{\vec{X}}(\Theta) = w_{\vec{Y}}(\Theta)$, and since a probability function is determined by its action on state descriptions, $W_{\vec{X}}$ and $w_{\vec{Y}}$ are equal as probability functions.

To see the converse does not hold, let L be the language containing a single binary relation symbol and let

$$\vec{Y} = \left\langle y^1_1 = \frac{3}{8}\ ,\ y^1_2 = \frac{5}{8}\ ;\ y^2_1 = \cdots = y^2_{10} = \frac{1}{4} \right\rangle .$$

Then

$$w_{\vec{Y}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{1}{4}$$

and

$$w_{\vec{Y}} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{1}{4},$$

however there is no $W_{\vec{X}}$ that assigns these same values to these state descriptions. □

The above theorem allows us to deduce that the $W_{\vec{X}}$ satisfy the independence principles we have previously encountered. See page 109 for a statement of the Strong Independence Principle, SIP, and page 65 for the Constant Independence Principle, IP.

Corollary 7.9. *The functions $W_{\vec{X}}$ satisfy SIP and hence also IP.*

Proof. By Theorem 7.8 every $W_{\vec{X}}$ is also one of the functions $w_{\vec{Y}}$. The $w_{\vec{Y}}$ satisfy SIP (and hence also IP) by Theorem 6.10, so the result follows. □

Finally, we show that the $W_{\vec{X}}$ can be used to make c_0^L and c_∞^L. $W_{\vec{X}}$ with $\vec{X}$ defined by

$$\left\langle x_1^s = x_2^s = \cdots = x_{2^{q_s}}^s = 2^{-q_s} \ : s \in \{1, \ldots, r\}, \ q_s \neq 0 \right\rangle$$

treats each $\pm R_d(a_{i_1}, \ldots, a_{i_{r_d}})$ as stochastically independent and each occurring with probability $\frac{1}{2}$. For c_0^L, we take the convex sum of all $\vec{X}$ where exactly one x_k^s is 1 for each arity s.

7.5 A Representation Theorem for e-Sgx

We prove that the $W_{\vec{X}}$ form the building blocks of all probability functions satisfying e-Sgx.

Theorem 7.10. *Let w be a probability function for an r-ary language L satisfying e-Sgx. Then there exists a normalised, σ-additive measure μ on the Borel subsets of $\mathbb{H}_L$ such that for any $\theta \in SL$*

$$w(\theta) = \int_{\mathbb{H}_L} W_{\vec{X}}(\theta) \ d\mu(\vec{X}). \tag{7.8}$$

Conversely, for a given measure μ on the Borel subsets of $\mathbb{H}_L$, the function defined by (7.8) *is a probability function on SL satisfying e-Sgx.*

Proof. We follow a similar method to the proof of de Finetti's Representation Theorem in the unary, see for example [36, Theorem 9.1], and to our proof of Theorem 5.9. Let w satisfy e-Sgx, let $u \in \mathbb{N}^+$ and let $\vec{U} = \langle u_1^s, \ldots, u_{2^{q_s}}^s \, : \, s \in \{1, \ldots, r\}$ and $q_s \neq 0 \rangle$ be an e-signature on u.

The number of state descriptions with this e-signature, denoted by $\mathcal{N}(\emptyset, \vec{U})$, is

$$\prod_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \binom{u^s}{u_1^s, \ldots, u_{2^{q_s}}^s}. \tag{7.9}$$

Let $\vec{M}$ be an e-signature, $m < u$ and let $\Theta(b_1, \ldots, b_m)$ be a state description with this e-signature. The number of state descriptions with e-signature $\vec{U}$ extending $\Theta(b_1, \ldots, b_m)$, denoted by $\mathcal{N}(\vec{M}, \vec{U})$, is

$$\prod_{\substack{s \in \{1,\ldots,r\} \\ q_s \neq 0}} \binom{u^s - m^s}{u_1^s - m_1^s \, , \, \ldots \, , \, u_{2^{q_s}}^s - m_{2^{q_s}}^s}. \tag{7.10}$$

We follow the convention that (7.10) is 0 if any of the terms are negative. We will write $w(\vec{M})$ for the value given by w to any state description with e-signature $\vec{M}$. This is valid since w satisfies e-Sgx. Let $eSign(u)$ denote the set containing all e-signatures $\vec{U}$ on u:

$$eSign(u) = \left\{ \vec{U} \, : \, \sum_{1 \leq k \leq 2^{q_s}} u_k^s = u^s \text{ for each } s \in \{1, \ldots, r\} \text{ with } q_s \neq 0 \right\}.$$

Using the fact that

$$w(\Theta(b_1, \ldots, b_m)) = \sum_{\Phi(b_1,\ldots,b_m,\ldots,b_u) \models \Theta(b_1,\ldots,b_m)} w(\Phi(b_1, \ldots, b_m, \ldots, b_u)),$$

we have

$$1 = w(\top) = \sum_{\vec{U} \in eSign(u)} \mathcal{N}(\emptyset, \vec{U}) \; w(\vec{U}), \tag{7.11}$$

$$w(\vec{M}) = \sum_{\vec{U} \in eSign(u)} \mathcal{N}(\vec{M}, \vec{U}) \; w(\vec{U}), \tag{7.12}$$

and hence

$$w(\vec{M}) = \sum_{\vec{U} \in eSign(u)} \frac{\mathcal{N}(\vec{M}, \vec{U})}{\mathcal{N}(\emptyset, \vec{U})} \, \mathcal{N}(\emptyset, \vec{U}) \; w(\vec{U}). \tag{7.13}$$

We will show that

$$\left| \left(\frac{\mathcal{N}(\vec{M},\vec{U})}{\mathcal{N}(\emptyset,\vec{U})} \right) - \prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{1\leq k\leq 2^{q_s}} \left(\frac{u_k^s}{u^s} \right)^{m_k^s} \right| \tag{7.14}$$

is of the order $\mathcal{O}(\sqrt{u}^{-1})$ and hence tends to 0 as $u \to \infty$ independently of the $u_1^s, \dots, u_{2^{q_s}}^s$. We follow the convention that if some $u_k^s = 0$ then terms involving it are missing from the product above.

Suppose first that $m_k^s < u_k^s$ for every k, s. The term $\left(\frac{\mathcal{N}(\vec{M},\vec{U})}{\mathcal{N}(\emptyset,\vec{U})}\right)$ in (7.14) can be written as

$$\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{1\leq k\leq 2^{q_s}} \left(\frac{u_k^s}{u^s} \right)^{m_k^s} \times \left(\frac{\prod_{1\leq k\leq 2^{q_s}} \left((1-(u_k^s)^{-1})\cdots(1-(m_k^s-1)\,(u_k^s)^{-1})\right)}{(1-u^{-s})\cdots(1-(m^s-1)u^{-s})} \right). \tag{7.15}$$

We will write P for the term

$$\left(\frac{\prod_{1\leq k\leq 2^{q_s}} \left((1-(u_k^s)^{-1})\cdots(1-(m_k^s-1)\,(u_k^s)^{-1})\right)}{(1-u^{-s})\cdots(1-(m^s-1)u^{-s})} \right).$$

If $m_k^s = 0$ then factors involving u_k^s do not appear in either term in (7.14), so we only need to consider those k for which $m_k^s > 0$. Consider the case that for some s, k with $m_k^s > 0$ we have $u_k^s < \sqrt{u^s}$. Then

$$\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{1\leq k\leq 2^{q_s}} \left(\frac{u_k^s}{u^s} \right)^{m_k^s} < \frac{1}{\sqrt{u^s}} \leq \frac{1}{\sqrt{u}}.$$

In addition, the other term in (7.15) is bounded by a constant independent of u and the u_k^s:

$$P < \left(\frac{1}{1-(m^s-1)m^{-s}} \right)^{m^s} = (m^s)^{m^s}.$$

So (7.15) is $\mathcal{O}(\sqrt{u}^{-1})$ and so (7.14) $\to 0$ as $u \to \infty$.

Now suppose that for every s, k with $m_k^s > 0$, $u_k^s > \sqrt{u^s}$. Notice that P can be written as the product of m^s fractions of the form $\frac{1-\alpha(u_k^s)^{-1}}{1-\beta u^{-s}}$ where $\alpha, \beta \in \{1, \dots, m^s\}$. Each such fraction is close to 1 since

$$\left| \frac{1-\alpha(u_k^s)^{-1}}{1-\beta u^{-s}} - 1 \right| = \left| \frac{\beta u^{-s} - \alpha(u_k^s)^{-1}}{1-\beta u^{-s}} \right| < 2\left(\beta u^{-s} + \alpha(u_k^s)^{-1}\right) < \frac{2}{\sqrt{u^s}}(\alpha+\beta) \leq \frac{4m^s}{\sqrt{u^s}},$$

where the inequalities hold since $\beta \leq m^s < \frac{u^s}{2}$ and so $1 - \beta u^{-s} > \frac{1}{2}$; because $u^{-s}, (u_k^s)^{-1} < \frac{1}{\sqrt{u^s}}$ and since $\alpha, \beta \leq m^s$, respectively. As $\frac{4m^s}{\sqrt{u^s}} \leq \frac{4m^s}{\sqrt{u}}$, the distance of each fraction from 1 is $\mathcal{O}(\sqrt{u}^{-1})$, so P is

$$\left(1 + \mathcal{O}(\sqrt{u}^{-1})\right)^{m^s} = 1 + \mathcal{O}\left(\sqrt{u}^{-1}\right).$$

Note that (7.14) can be written as

$$\left| \prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{1\leq k\leq 2^{q_s}} \left(\frac{u_k^s}{u^s}\right)^{m_k^s} (1-P) \right| \tag{7.16}$$

and that the product $\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{1\leq k\leq 2^{q_s}} \left(\frac{u_k^s}{u^s}\right)^{m_k^s}$ is bounded by 1. So (7.16) is $\mathcal{O}(\sqrt{u}^{-1})$ and thus so is (7.14). So we have shown that (7.14) $\to 0$ as $u \to \infty$ when $m_k^s < u_k^s$ for every s, k.

Suppose $u_k^s < m_k^s$ for some k and s. Then $\frac{\mathcal{N}(\vec{M},\vec{U})}{\mathcal{N}(\emptyset,\vec{U})} = 0$ since no state description with signature $\vec{U}$ can extend a state description with signature $\vec{M}$. In addition, $m_k^s > 0$ and $u_k^s < \sqrt{u^s}$, so we would have

$$\prod_{\substack{s\in\{1,\dots,r\}\\ q_s\neq 0}} \prod_{1\leq k\leq 2^{q_s}} \left(\frac{u_k^s}{u^s}\right)^{m_k^s} < \frac{1}{\sqrt{u^s}} \leq \frac{1}{\sqrt{u}}$$

and consequently, arguing as above, (7.14) $\to 0$ as $u \to \infty$ in this final case too.

Define $\vec{X}_{\vec{U}}$ by

$$x_k^s = \frac{u_k^s}{u^s}$$

for $k \in \{1, \dots, 2^{q_s}\}$ and $s \in \{1, \dots, r\}$, $q_s \neq 0$. The proof is completed similarly to the proof of Theorem 5.9. □

Theorem 7.11. *Let w be a probability function on SL. Then w satisfies e-Sgx and IP if and only if $w = W_{\vec{X}}$ for some $\vec{X} \in \mathbb{H}_L$.*

Proof. The proof is the same as that of Theorem 5.10. □

7.6 Element-based Instantial Relevance

Our definition of instantial relevance in the atom-based theory involved the agent acquiring new information in the form of a g-atom instantiated by some (possibly new) individuals. In our current context, we will define instantial relevance with the additional information being an instantiated s-ary element.

Polyadic Principle of Instantial Relevance (elements version), e-PPIR
Let $\Delta(a_1, \ldots, a_m)$ be an e-partial state description, $s \in \{1, \ldots, r\}$ with $q_s \neq 0$, and let $\delta^s_k(x_1, \ldots, x_s)$ be an s-ary element. Let $b_1, \ldots, b_s$, $b'_1, \ldots, b'_s$ be such that

$$\Delta \wedge \delta^s_k(b_1, \ldots, b_s) \wedge \delta^s_k(b'_1, \ldots, b'_s)$$

is consistent. Then

$$w(\delta^s_k(b_1, \ldots, b_s) \mid \Delta) \leq w(\delta^s_k(b_1, \ldots, b_s) \mid \delta^s_k(b'_1, \ldots, b'_s) \wedge \Delta). \tag{7.17}$$

Notice that if $\Delta \models \delta^s_k(b'_1, \ldots, b'_s)$ or $\Delta \models \delta^s_k(b_1, \ldots, b_s)$ then (7.17) holds with equality.

Theorem 7.12. *Let w be a probability function satisfying e-Sgx. Then w satisfies e-PPIR.*

Proof. The theorem is proved similarly to Theorem 5.13. $\square$

Chapter 8

Homomorphisms and Degeneracy

8.1 Introduction

We introduce the concepts of *homomorphisms* and *degeneracy* in Polyadic Inductive Logic. The results in this chapter are somewhat preliminary, and lay the groundwork for future work in this area.

The idea is as follows. As usual, an agent inhabits a structure $\mathcal{M}$ for a language L, but does not know which of the L-structures $\mathcal{M}$ is. The agent's aim is to choose a rational probability function w that assigns degrees of belief to sentences of L. If the agent were able to represent w by a *simpler* probability function that assigns the same values as w to sentences of L, it would surely be rational of them to do so. An obvious interpretation of 'simpler' here is that w is represented by a probability function on a language of lower arity. One could argue, for example, that the agent's capacity for holding information, their 'memory', is a limited quantity, and thus an agent might only 'remember' the first n coordinates of an r-ary relation symbol when $r > n$. Though this interpretation is by no means the only one, and we leave the precise definition of what simpler might mean open to future investigations.

Aside from the seeming rationality of such a concept, the other main advantage is a practical one. We have already seen that the mathematics involved gets considerably more complicated as the arity of the language increases. We have a better understanding of the behaviour of probability functions on lower arity languages, and unary and binary languages in particular are easier to visualise. We have demonstrated a use for such an approach when we proved that INV does not imply SReg in Section 2.3, where we used a result on unary languages to deduce it for non-unary languages.

Notation

We use the following notation in what follows. For a language L, let $\mathcal{T}L$ again denote the set of structures $\mathcal{M}$ for L as defined on page 11. Recall that the universe of $\mathcal{M}$, which we will denote by $|\mathcal{M}|$, is $\{a_i \mid i \in \mathbb{N}^+\}$ (with each a_i interpreted as itself). We will use the fact[1] that if $\theta \in SL$ is consistent then there is a structure $\mathcal{M} \in \mathcal{T}L$ such that $\mathcal{M} \models \theta$. Note that this means also that if there is an L-structure in which θ does not hold, there must be an $\mathcal{M} \in \mathcal{T}L$ such that $\mathcal{M} \models \neg\theta$, since $\neg\theta$ must be consistent.

Let FL be the set of formulae of L and $\phi \in FL$. We denote by $Var(\phi)$ the set of variables appearing in ϕ (including bound variables) and by $FrVar(\phi)$ the set of free variables mentioned in ϕ. $\phi(x_1, \ldots, x_n)$ is taken to imply that $FrVar(\phi) \subseteq \{x_1, \ldots, x_n\}$ (so that x_i, $i \in \{1, \ldots, n\}$ may or may not appear in ϕ). We assume throughout this chapter that the languages we deal with are non-empty.

8.2 Homomorphisms

Definition 8.1. For languages L, L' we define $\kappa : FL' \to FL$ to be a *homomorphism from L' to L* if and only if κ satisfies the following conditions:

For R an r-ary relation symbol of L' and terms $t_1, \ldots, t_r$ (so that $R(t_1, \ldots, t_r)$ is an atomic formula of L')

(i) $\kappa(R(t_1, \ldots, t_r)) = \kappa(R(x_1, \ldots, x_r))(t_1, \ldots, t_r)$.

(ii) $FrVar(\kappa(R(t_1, \ldots, t_r))) \subseteq FrVar(R(t_1, \ldots, t_r))$.

For $\theta, \phi, \exists x_j\, \psi(x_j) \in FL'$,

(iii) $\kappa(\neg\theta) = \neg\kappa\theta$.

(iv) $\kappa(\theta \vee \phi) = \kappa\theta \vee \kappa\phi$.

[1]See [36, Chapter 2] for details.

(v) $\kappa(\exists x_j\, \psi(x_j)) = \exists x_j\, \kappa(\psi(x_j))$.

Notice that (i) enables us to define the action of κ on atomic sentences of L', and that (ii) ensures that κ introduces no new variables - a restriction on the 'complexity' of κ. In what follows $\kappa\theta$ will sometimes be used instead of $\kappa(\theta)$ to aid readability.

Let $\kappa : FL' \to FL$ be a homomorphism from L' to L and let $\mathcal{M} \in \mathcal{T}L$. Define $\mathcal{M}' \in \mathcal{T}L'$ with the same universe as $\mathcal{M}$ such that for any atomic sentence[2] θ of L'

$$\mathcal{M}' \models \theta \iff \mathcal{M} \models \kappa\theta. \tag{8.1}$$

Proposition 8.2. *Let κ be a homomorphism from L' to L and let $\mathcal{M}, \mathcal{M}'$ be the structures for L, L' respectively, defined by* (8.1)*. Then for any sentence ϕ of L'*

$$\mathcal{M}' \models \phi \iff \mathcal{M} \models \kappa\phi. \tag{8.2}$$

Proof. We prove the proposition by induction on the complexity of sentences of L'. (8.2) holds for atomic sentences of L' by the definition (8.1) of the structures $\mathcal{M}, \mathcal{M}'$. Suppose $\theta, \phi, \psi(a_i)$ are sentences of L' for which (8.2) holds.

Consider the L' sentence $\neg\theta$. Since by our assumption (8.2) holds for θ, we have that $\mathcal{M}' \not\models \theta \iff \mathcal{M} \not\models \kappa\theta$. However then

$$\mathcal{M}' \models \neg\theta \iff \mathcal{M}' \not\models \theta \iff \mathcal{M} \not\models \kappa\theta \iff \mathcal{M} \models \neg\kappa\theta$$

using the definition of interpretation of sentences for L' and L.

Consider next the sentence $(\theta \vee \phi) \in SL'$. We have

$$\begin{aligned} \mathcal{M}' \models \theta \vee \phi &\iff \mathcal{M}' \models \theta \text{ or } \mathcal{M}' \models \phi \\ &\iff \mathcal{M} \models \kappa\theta \text{ or } \mathcal{M} \models \kappa\phi \\ &\iff \mathcal{M} \models (\kappa\theta \vee \kappa\phi) \\ &\iff \mathcal{M} \models \kappa(\theta \vee \phi), \end{aligned}$$

where the first and third bi-implications follow from the interpretation of L'-sentences and L-sentences respectively; the second bi-implication results from our assumption

[2] so θ has the form $R(a_{i_1}, \ldots, a_{i_r})$ for some relation symbol R of L' (of arity r).

that (8.2) holds for θ, ϕ, and the final bi-implication from property (iv) of κ. At this point we have shown by induction that (8.2) holds for all quantifier-free sentences of L'.

Finally, consider $\exists x_j\, \psi(x_j) \in SL'$.

$$\begin{aligned}
\mathcal{M}' \models \exists x_j\, \psi(x_j) &\iff \mathcal{M}' \models \psi(a_i) \text{ for some } a_i \in |\mathcal{M}'| \\
&\iff \mathcal{M} \models \kappa(\psi(a_i)) \text{ for (the same) } a_i \in |\mathcal{M}| \\
&\iff \mathcal{M} \models \exists x_j\, \kappa(\psi(x_j)) \\
&\iff \mathcal{M} \models \kappa(\exists x_j\, \psi(x_j))
\end{aligned}$$

by the interpretation of L'-sentences, by the inductive hypothesis for $\psi(a_i)$ and since $|\mathcal{M}'| = |\mathcal{M}|$ by definition of these structures, by the interpretation of L-sentences, and by property (v) of κ respectively.

Since every L'-sentence is either an atomic-sentence, or follows by a finite number of steps from the three inductive cases, we conclude that (8.2) holds for any $\phi \in SL'$. □

Corollary 8.3. *For $\theta, \phi \in SL'$ and κ a homomorphism from L' to L,*

$$\theta \equiv \phi \implies \kappa\theta \equiv \kappa\phi.$$

Proof. Suppose $\theta \equiv \phi$. Then exactly the same L'-structures satisfy both θ and ϕ. In particular, for every L'-structure of the form $\mathcal{M}'$ as defined by (8.1),

$$\mathcal{M}' \models \theta \iff \mathcal{M}' \models \phi.$$

From Proposition 8.2 we have that $\mathcal{M}' \models \theta \iff \mathcal{M} \models \kappa\theta$ and similarly for ϕ. Therefore,

$$\begin{aligned}
\theta \equiv \phi &\implies (\mathcal{M}' \models \theta \iff \mathcal{M}' \models \phi) \\
&\iff (\mathcal{M} \models \kappa\theta \iff \mathcal{M} \models \kappa\phi) \\
&\iff \kappa\theta \equiv \kappa\phi
\end{aligned}$$

using our remark from page 137 for the final bi-implication, since if every $\mathcal{M} \in \mathcal{T}L$ is such that $\mathcal{M} \models \kappa\theta \iff \mathcal{M} \models \kappa\phi$, then this holds for every L-structure. □

As a consequence of the above corollary, when dealing with homomorphisms from a language L' to a language L, we may treat sentences of L' up to logical equivalence only.

8.3 Degenerate Probability Functions

Let w be a probability function on L and let κ be a homomorphism from L' to L. Define $w' : SL' \to [0,1]$ by

$$w'(\theta) = w(\kappa\theta). \tag{8.3}$$

Proposition 8.4. *w' defined by* (8.3) *is a probability function on L'. Moreover, if w satisfies Ex, than so does w'.*

Proof. We show that w' satisfies (P1)-(P3) from page 12. Let $\mathcal{M}' \in \mathcal{T}L'$ be as in (8.1), so by Proposition 8.2 for all $\theta \in SL'$ and $\mathcal{M} \in \mathcal{T}L$ we have

$$\mathcal{M}' \models \theta \iff \mathcal{M} \models \kappa\theta.$$

Let $\theta \in SL'$ be such that $\models \theta$. Then in particular, for every L'-structure of the form $\mathcal{M}'$, $\mathcal{M}' \models \theta$, and hence for every $\mathcal{M} \in \mathcal{T}L$, $\mathcal{M} \models \kappa\theta$. Therefore, following similar reasoning to above, this holds for *all* L-structures, and so $\models \kappa\theta$. Now $w'(\theta) = w(\kappa\theta) = 1$ since w is a probability function so it satisfies (P1). Thus w' satisfies (P1).

Suppose $\theta \models \neg\phi$. This is logically equivalent to $\models \theta \to \neg\phi$, and hence by (P1) above we have

$$\begin{aligned}
\models \theta \to \neg\phi &\Longrightarrow \models \kappa(\theta \to \neg\phi) \\
&\iff \models \kappa(\neg\theta \vee \neg\phi) \\
&\iff \models \kappa(\neg\theta) \vee \kappa(\neg\phi) \\
&\iff \models \neg\kappa\theta \vee \neg\kappa\phi \\
&\iff \models \kappa\theta \to \neg\kappa\phi \\
&\iff \ \kappa\theta \models \neg\kappa\phi.
\end{aligned}$$

where the bi-implications are obtained by logical equivalence and by the properties of κ. Now, $w'(\theta \vee \phi) = w(\kappa(\theta \vee \phi)) = w(\kappa\theta \vee \kappa\phi)$. Since w satisfies (P2) and $\kappa\theta \models \neg\kappa\phi$, $w(\kappa\theta \vee \kappa\phi) = w(\kappa\theta) + w(\kappa\phi)$. But $w(\kappa\theta) = w'(\theta)$ and $w(\kappa\phi) = w'(\phi)$. So $w'(\theta \vee \phi) = w'(\theta) + w'(\phi)$, and hence w' satisfies (P2).

To see that (P3) holds too, notice that

$$\begin{aligned}
w'(\exists x_j\, \psi(x_j)) &= w(\kappa(\exists x_j\, \psi(x_j))) \\
&= w(\exists x_j\, \kappa(\psi(x_j))) \\
&= \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \kappa(\psi(a_i))\right) \\
&= \lim_{n\to\infty} w\left(\kappa\left(\bigvee_{i=1}^{n} \psi(a_i)\right)\right) \\
&= \lim_{n\to\infty} w'\left(\bigvee_{i=1}^{n} \psi(a_i)\right),
\end{aligned}$$

where the third equality follows since w satisfies (P3), and the other equalities obtain from the properties of a homomorphism and the definition of w'. Therefore, w' satisfies conditions (P1)-(P3), so w' is a probability function on SL'.

In order to see that if w satisfies Ex than so does w', observe that for any $\theta \in FL'$ we have

$$\kappa(\theta(t_1,\ldots,t_n)) = \kappa(\theta(x_1,\ldots,x_n))(t_1,\ldots,t_n).$$

This holds for atomic formulae by property (i) of Definition 8.1, and follows by induction on formula complexity for all formulae of L'. For example, suppose it holds for $\theta(t_1,\ldots,t_n) \in FL'$. Then

$$\begin{aligned}
\kappa(\neg\theta(t_1,\ldots,t_n)) &= \neg\kappa(\theta(t_1,\ldots,t_n)) \\
&= \neg\kappa(\theta(x_1,\ldots,x_n))(t_1,\ldots,t_n) = \kappa(\neg\theta(x_1,\ldots,x_n))(t_1,\ldots,t_n).
\end{aligned}$$

The other cases are similar. Suppose $\kappa\theta$ is the formula $\phi \in FL$. Then by the above observation,

$$\kappa(\theta(a_1,\ldots,a_n)) = \kappa(\theta(x_1,\ldots,x_n))(a_1,\ldots,a_n) = \phi(a_1,\ldots,a_n),$$

$$\kappa(\theta(b_1,\ldots,b_n)) = \kappa(\theta(x_1,\ldots,x_n))(b_1,\ldots,b_n) = \phi(b_1,\ldots,b_n).$$

Since w satisfies Ex, we have

$$w(\kappa(\theta(a_1,\ldots,a_n))) = w(\phi(a_1,\ldots,a_n)) = w(\phi(b_1,\ldots,b_n)) = w(\kappa(\theta(b_1,\ldots,b_n))).$$

So $w'(\theta(a_1,\ldots,a_n)) = w'(\theta(b_1,\ldots,b_n))$, and thus w' satisfies Ex. □

We say a probability function w' on SL' is (L',L)-*degenerate* if for some simpler[3] language L, there exists a homomorphism $\kappa : FL' \to FL$ and a probability function w on SL such that

$$w'(\theta) = w(\kappa(\theta)). \tag{8.4}$$

w' is *degenerate* if it is (L',L)-degenerate for some L.

Proposition 8.4 enables us to express an (L',L)-degenerate probability function in terms of a probability function on SL using the homomorphism κ. It also tells us that if the probability function on SL satisfies Ex, then an (L',L)-degenerate probability function expressed by (8.4) will also satisfy Ex. On the other hand, degenerate probability functions do not necessarily preserve other properties, as we show next.

Proposition 8.5. *Let w' be a degenerate probability function as in* (8.4). *Then w' may fail Vx, PIP, Sx, Px, Reg and SN even if w satisfies these principles.*

Proof. We prove the proposition by providing examples of probability functions w' as in (8.4) where the above principles are satisfied by w but not by w'. Let L_R be a language containing a binary relation symbol R (possibly amongst others) and let L_P be the same language with R replaced by a unary predicate symbol P (where $P \notin L_R$).

For $\psi \in SL_R$, let $\kappa\psi \in SL_P$ denote the result of replacing each occurrence of $R(t_1,t_2)$ in ψ, where t_1,t_2 are any terms, by $P(t_1)$ and keeping all other relation symbols (if any) unchanged. Then $\kappa : FL_R \to FL_P$ defines a homomorphism from L_R to L_P since it satisfies (i)-(v) of Definition 8.1.

For $\mathcal{M} \in \mathcal{T}L_P$, define $\mathcal{M}' \in \mathcal{T}L_R$ (with the same universe as $\mathcal{M}$) to interpret all symbols in $L_P \cap L_R$ in the same way as $\mathcal{M}$, and be such that

$$\mathcal{M}' \models R(t_1,t_2) \iff \mathcal{M} \models P(t_1).$$

[3]where, as indicated at the beginning of the chapter, we adopt an intuitive interpretation of 'simpler' suitable for our purposes.

Define $w_R : SL_R \to [0, 1]$ by

$$w_R(\theta) = c_\infty^{L_P}(\kappa\theta).$$

Then w_R is a probability function on SL_R by Proposition 8.4, and $c_\infty^{L_P}$ is a probability function on SL_P that satisfies Vx (on non-unary languages), PIP, Sx, Px and SN.

Consider the sentences of L_R

$$\theta = (R(a_1, a_1) \wedge R(a_1, a_2)), \tag{8.5}$$

$$\phi = (R(a_1, a_1) \wedge R(a_2, a_1)). \tag{8.6}$$

If w_R satisfied Vx, we should get that $w_R(\theta) = w_R(\phi)$ and by the definition of w_R and κ, we would thus require[4] $c_\infty^{L_P}(P(a_1)) = c_\infty^{L_P}(P(a_1) \wedge P(a_2))$. However $c_\infty^{L_P}(P(a_1)) = 1/2$ and $c_\infty^{L_P}(P(a_1) \wedge P(a_2)) = 1/4$, thus Vx fails for w_R.

We now address PIP. Vx is implied by PIP, therefore since Vx does not hold, w_R cannot satisfy PIP. To see it directly, let $\Theta(a_1, a_2, a_3)$, $\Phi(a_1, a_2, a_3)$ be the similar state descriptions represented respectively by the matrices

$$\begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{matrix} \qquad\qquad \begin{matrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{matrix}\ .$$

We have

$$\kappa\Theta = P(a_1) \wedge \neg P(a_2) \wedge P(a_3),$$
$$\kappa\Phi = \neg P(a_1) \wedge P(a_1) \wedge P(a_2) \wedge \neg P(a_3) \wedge P(a_3),$$

and $c_\infty^{L_P}(\kappa\Theta) = 1/8$ while $c_\infty^{L_P}(\kappa\Phi) = 0$, so PIP does not hold for w_R. This means that Sx fails for w_R too, since (Sx implies PIP and) the above state descriptions have the same spectrum, $\{2, 1\}$.

To see that Px fails, suppose that L_R contains (at least) one other binary relation symbol Q (so Q is also in L_P). For $\psi \in SL_R$, let $\psi^* \in SL_R$ denote the sentence resulting by simultaneously swapping each occurrence of the relation symbol R by Q and Q by R. Let θ be as in (8.5), so θ^* is $(Q(a_1, a_1) \wedge Q(a_1, a_2))$. Then we have

$$\kappa\theta = P(a_1)$$

[4] As usual, we identify logically equivalent sentences since they are given the same probability by any probability function.

$$\kappa\theta^* = Q(a_1, a_1) \wedge Q(a_1, a_2),$$

and $c_\infty^{L_P}(\kappa\theta) = 1/2$ while $c_\infty^{L_P}(\kappa\theta^*) = 1/4$. So $w_R(\theta) \neq w_R(\theta^*)$, and hence w_R does not satisfy Px.

Regularity fails for w_R, and also when w_R is defined with $c_\infty^{L_P}$ replaced by *any* other probability function w_P on SL_P, since

$$w_R(R(a_1, a_1) \wedge \neg R(a_1, a_2)) = w_P(P(a_1) \wedge \neg P(a_1)) = 0$$

but $R(a_1, a_1) \wedge \neg R(a_1, a_2)$ is satisfiable.

Finally, to show that SN is not preserved by degenerate probability functions, let L_R be as before and let $L_{P,Q}$ be the same language with R replaced by two unary predicate symbols P and Q (where $P, Q \notin L_R$). For $\psi \in SL_R$, let $\lambda\psi \in SL_{P,Q}$ denote the result of replacing each occurrence of $R(t_1, t_2)$ in ψ by $P(t_1) \wedge Q(t_2)$. λ satisfies (i)-(v) of Definition 8.1, so λ is a homomorphism from L_R to $L_{P,Q}$. Define the probability function $v_R : SL_R \to [0, 1]$ by

$$v_R(\psi) = c_\infty^{L_{P,Q}}(\lambda\psi).$$

For v_R to satisfy SN we must have, for example, $v_R(R(a_1, a_2)) = v_R(\neg R(a_1, a_2))$. However

$$\begin{aligned} v_R(R(a_1, a_2)) &= c_\infty^{L_{P,Q}}(P(a_1) \wedge Q(a_2)) = 1/4 \\ v_R(\neg R(a_1, a_2)) &= c_\infty^{L_{P,Q}}(\neg\,(P(a_1) \wedge Q(a_2))) = c_\infty^{L_{P,Q}}(\neg P(a_1) \vee \neg Q(a_2)) = 3/4 \end{aligned}$$

and so SN fails. □

We conclude the chapter by showing that using a homomorphism, any probability function on an r-ary language can be represented by a probability function on a binary language[5].

To simplify notation, let L be the language containing a single r-ary relation symbol R.[6] Let L_r be the language with binary relation symbols $B_1, \ldots, B_r$. For $\theta \in SL$, let

[5]This result is based on a suggestion which arose during a joint discussion with the University of Manchester Inductive Logic Group.

[6]The result generalises to any r-ary language using a similar method.

$\kappa\theta \in SL_r$ be the result of replacing $R(t_1, \ldots, t_r)$ everywhere in θ by

$$\exists x\, (B_1(x, t_1) \wedge B_2(x, t_2) \wedge \cdots \wedge B_r(x, t_r)).$$

Then $\kappa : FL \to FL_r$ defines a homomorphism from L to L_r: Properties (i) and (ii) of Definition 8.1 clearly hold. To see that κ satisfies the other properties, notice that they hold when $\theta, \phi, \psi(x_j)$ are atomic, and hence also when $\theta, \phi, \psi(x_j)$ are any L-formulae by induction on the complexity of L-formulae.

Let σ be a bijective map from $\mathbb{N}^+$ to $(\mathbb{N}^+)^r$ and for $\mathcal{M} \in \mathcal{T}L$ let $\mathcal{M}^* \in \mathcal{T}L_r$ be the structure for L_r such that for $\sigma(s) = \langle i_1, \ldots, i_r \rangle$

$$\mathcal{M}^* \models B_1(a_s, a_{i_1}) \wedge B_2(a_s, a_{i_2}) \wedge \cdots \wedge B_r(a_s, a_{i_r}) \iff \mathcal{M} \models R(a_{i_1}, \ldots, a_{i_r}).$$

Then

$$\mathcal{M} \models R(a_{i_1}, \ldots, a_{i_r}) \iff \mathcal{M}^* \models \exists x\, (B_1(x, a_{i_1}) \wedge B_2(x, a_{i_2}) \wedge \cdots \wedge B_r(x, a_{i_r})),$$

and hence by Proposition 8.2, for any $\theta \in SL$

$$\mathcal{M} \models \theta \iff \mathcal{M}^* \models \kappa\theta.$$

For $\mathcal{M} \in \mathcal{T}L$, let $V_\mathcal{M} : SL \to \{0, 1\}$ be the probability function[7] defined by

$$V_\mathcal{M}(\theta) = \begin{cases} 1 & \text{if } \mathcal{M} \models \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Let w be a probability function on SL. By [36, Corollary 7.2], $w(\theta)$ can be represented by

$$w(\theta) = \int_{\mathcal{T}L} V_\mathcal{M}(\theta)\; d\mu(\mathcal{M})$$

for μ a σ-additive, normalised measure on the σ-algebra of subsets of $\mathcal{T}L$ generated by the subsets[8] $[\theta] = \{\mathcal{M} \in \mathcal{T}L \mid \mathcal{M} \models \theta\}$ for $\theta \in SL$.

Define $v : SL_r \to [0, 1]$ by

$$v(\phi) = \int_{\mathcal{T}L} V_{\mathcal{M}^*}(\phi)\; d\mu(\mathcal{M}).$$

[7] from [36, Chapter 3].

[8] as defined on page 18.

Notice that this definition is valid, since for $\phi \in SL_r$ the set $\{\mathcal{M} \in \mathcal{T}L \mid \mathcal{M}^* \models \phi\}$ is Borel and hence measurable, and furthermore v indeed defines a probability function on SL_r since (P1) and (P2) clearly hold, and (P3) holds by Lebesgue's Dominated Convergence Theorem.

So for $\theta \in SL$ we have

$$w(\theta) = \int_{\mathcal{T}L} V_{\mathcal{M}}(\theta)\ d\mu(\mathcal{M}) = \int_{\mathcal{T}L} V_{\mathcal{M}^*}(\kappa\theta)\ d\mu(\mathcal{M}) = v(\kappa\theta).$$

That is, the probability function w on SL can be represented by the probability function v on SL_r so w is (L, L_r)-degenerate.

Chapter 9

Conclusions

Our aim for this thesis has been to investigate rational principles based on symmetry in Polyadic Inductive Logic. We have done this by furthering our understanding of existing symmetry principles - focusing on the recently introduced Permutation Invariance Principle, as well as presenting new possible avenues.

We began by investigating properties of PIP. We tested its consistency with longer standing principles of Inductive Logic, and found that PIP is determined by a finite set of permutations acting on a finite set of formulae. This allowed us to extend the key unary notion of atoms to polyadic languages.

Following this, we explored the relationship between PIP, Spectrum Exchangeability and Language Invariance, with Sx and Li currently the most studied of polyadic rational principles. This helped us understand the behaviour of PIP and its standing within polyadic symmetry. A desirable extension to this investigation would be to generalise the functions $u_{\bar{E}}^{\bar{p},L}$ to any $\bar{p} \in \mathbb{B}$, and eventually to characterise the probability functions satisfying PIP via a representation theorem.

Using our definition of polyadic atoms, we then proceeded to investigate PIP as a generalisation of the popular unary principle of Atom Exchangeability. We found that it is in fact PIP, rather than the previously thought Sx, that stands as the natural polyadic extension of Ax. We concluded this part of the investigation with a probe into generalising Ax under the assumption of Ex and proposed the principle of Atom-based Spectrum Exchangeability. This is a direction still worthy of further thought in our opinion.

Inspired by new polyadic generalisations of the unary notion of a spectrum, we began exploring its close neighbour, the signature. The unary version of Ex expressed in terms of invariance under signatures formed the starting point for the second main line of investigation in this thesis. We presented a polyadic formulation of a signature, Signature Exchangeability and the Strong Independence Principle which arose from it, and a polyadic principle of instantial relevance. To our knowledge, this offers the first polyadic generalisation of the signature-based unary theory. In addition, in the binary case we were able to provide a complete characterisation of the functions satisfying Signature Exchangeability. Further research should be dedicated to checking whether it is possible to extend the representation to all polyadic languages.

We also presented an alternative, simpler yet disjointed, approach to these ideas, based on elements instead of atoms as the building blocks of polyadic languages. This approach was closer in spirit to the unary theory, but arguably captured the essence of Polyadic Inductive Logic less successfully.

From a broader perspective, this investigation and in particular the introduction of atoms as the polyadic building blocks, opens the door to generalising more results from Unary Inductive Logic which currently have no polyadic counterpart. We have mentioned there have already been results generalising Johnson's Sufficientness Postulate in terms of atoms [40, 42]. New possibilities could be extending the Unary Principle of Induction (see page 65) and the Only Rule [36, Theorem 21.6] along these lines.

Finally, we introduced the concepts of a homomorphism and a degenerate probability function into Inductive Logic. While preliminary, we believe this is a fruitful direction for future research, both in terms of its purported rationality and for its practical / mathematical applications. This investigation lays the foundations for any such work.

We have shed some light on a significant portion of polyadic symmetry. As always, new knowledge leads to new questions, and Polyadic Inductive Logic being a relatively young area of research means there is much yet to be discovered.

Bibliography

[1] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999.

[2] R. Carnap. On inductive logic. *Philosophy of Science*, 12(2):72–97, 1945.

[3] R. Carnap. *Logical Foundations of Probability*. The University of Chicago Press, Chicago, Ill., 1950.

[4] R. Carnap. *The Continuum of Inductive Methods*. The University of Chicago Press, Chicago, Ill., 1952.

[5] R. Carnap. The aim of inductive logic. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science*, pages 303 – 318. Stanford University Press, Stanford, Calif., 1962.

[6] R. Carnap. Replies and systematic expositions. In P. A. Schilpp, editor, *The Philosophy of Rudolf Carnap*, pages 859–1016. Open Court Publishing Company, La Salle, Ill., 1963.

[7] R. Carnap. A basic system of inductive logic. In R. Carnap and R.C. Jeffrey, editors, *Studies in Inductive Logic and Probability*, volume I, pages 33–165. University of California Press, Berkeley, Calif., 1971.

[8] R. Carnap. Notes on probability and induction. In J. Hintikka, editor, *Rudolf Carnap, Logical Empiricist*, pages 293–324. D. Reidel Publishing Company, Dordrecht-Boston, Mass., 1975.

[9] R. Carnap. A basic system of inductive logic. In R.C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume II, pages 7–155. University of California Press, Berkeley, Calif., 1980.

[10] N. J. Cutland. Loeb measure theory. In N. J. Cutland, F. Oliveira, V. Neves, and J. Sousa-Pinto, editors, *Developments in Nonstandard Mathematics*, volume 336 of *Pitman Research Notes in Mathematics Series*, pages 151–177. Longman Press, Longman, Harlow, 1995.

[11] B. de Finetti. Sul significato soggettivo dell probabilità. *Fundamenta Mathematicae*, 17:298–329, 1931.

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

[13] H. Gaifman. Concerning measures in first order calculi. *Israel Journal of Mathematics*, 2:1–18, 1964.

[14] A. Hill and J. B. Paris. An analogy principle in inductive logic. *Annals of Pure and Applied Logic*, 164(12):1293–1321, 2013.

[15] A. Hill and J. B. Paris. The counterpart principle of analogical support by structural similarity. *Erkenntnis*, 79(6):1169–1184, 2014.

[16] D. N. Hoover. Relations on probability spaces and arrays of random variables. Preprint, Institute of Advances Study, Princeton, 1979.

[17] W. E. Johnson. Probability: The deductive and inductive problems. *Mind*, 41:409–423, 1932.

[18] J. M. Keynes. *A Treatise on Probability.* Macmillan & Co., 1921.

[19] P. H. Krauss. Representation of symmetric probability models. *The Journal of Symbolic Logic*, 34(2):183–193, 1969.

[20] J. Landes. *The Principle of Spectrum Exchangeability within Inductive Logic.* PhD thesis, University of Manchester, 2009. Available at `http://www.maths.manchester.ac.uk/~jeff/`.

[21] J. Landes, J. B. Paris, and A. Vencovská. Language invariance and spectrum exchangeability in inductive logic. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 4724 of *Lecture Notes in Computer Science*, pages 151–160. Springer Berlin Heidelberg, 2007.

[22] J. Landes, J. B. Paris, and A. Vencovská. Some aspects of polyadic inductive logic. *Studia Logica*, 90(1):3–16, 2008.

[23] J. Landes, J. B. Paris, and A. Vencovská. Representation theorems for probability functions satisfying spectrum exchangeability in inductive logic. *International Journal of Approximate Reasoning*, 51(1):35–55, 2009.

[24] J. Landes, J. B. Paris, and A. Vencovská. A characterization of the language invariant families satisfying spectrum exchangeability in polyadic inductive logic. *Annals of Pure and Applied Logic*, 161(6):800–811, 2010.

[25] J. Landes, J. B. Paris, and A. Vencovská. A survey of some recent results on spectrum exchangeability in polyadic inductive logic. *Synthese*, 181(1):19–47, 2011.

[26] P. A. Loeb. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, 211:113–122, 1975.

[27] P. Maher. Probabilities for two properties. *Erkenntnis*, 52(1):63–91, 2000.

[28] P. Maher. Probabilities for multiple properties: the models of Hesse and Carnap and Kemeny. *Erkenntnis*, 55(2):183–216, 2001.

[29] C. J. Nix. *Probabilistic Induction in the Predicate Calculus*. PhD thesis, University of Manchester, 2005. Available at `http://www.maths.manchester.ac.uk/~jeff/`.

[30] C. J. Nix and J. B. Paris. A continuum of inductive methods arising from a generalized principle of instantial relevance. *Journal of Philosophical Logic*, 35(1):83–115, 2006.

[31] C. J. Nix and J. B. Paris. A note on binary inductive logic. *Journal of Philosophical Logic*, 36(6):735–771, 2007.

[32] J. B. Paris and A. Vencovská. From unary to binary inductive logic. In J. van Benthem, A. Gupta, and E. Pacuit, editors, *Games, Norms and Reasons*, volume 353 of *Synthese Library*, pages 199–213. Springer, Dordrecht, 2011.

[33] J. B. Paris and A. Vencovská. A note on Nathanial's invariance principle in polyadic inductive logic. In *Logic and its Applications*, volume 6521 of *Lecture Notes in Computer Science*, pages 137–146. Springer, Berlin, 2011.

[34] J. B. Paris and A. Vencovská. Symmetry's end? *Erkenntnis*, 74(1):53–67, 2011.

[35] J. B. Paris and A. Vencovská. Symmetry in polyadic inductive logic. *Journal of Logic, Language and Information*, 21:189–216, 2012.

[36] J. B. Paris and A. Vencovská. *Pure Inductive Logic*. In the Association of Symbolic Logic series Perspectives in Mathematical Logic. Cambridge University Press, 2015.

[37] J.B. Paris and P. Waterhouse. Atom exchangeability and instantial relevance. *Journal of Philosophical Logic*, 38(3):313–332, 2009.

[38] F. P. Ramsey. *The Foundations of Mathematics and other Logical Essays*. Routledge & Kegan Paul, London, 1931.

[39] T. Ronel and A. Vencovská. Invariance principles in polyadic inductive logic. *Logique et Analyse*, 228:541–561, 2014.

[40] T. Ronel and A. Vencovská. The principle of signature exchangeability. To appear in *Journal of Applied Logic*, 2015.

[41] H. L. Royden. *Real Analysis*. Macmillan Publishing Co., New York, second edition, 1968.

[42] A. Vencovská. Extending Carnap's continuum to binary relations. In *Logic and its Applications*, volume 8923 of *Lecture Notes in Computer Science*, pages 207–217. Springer, 2015.

[43] S. L. Zabell. *Symmetry and its Discontents*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, New York, 2005.

[44] S. L. Zabell. Carnap and the logic of induction. In D. M. Gabbay, S. Hartman, and J. Woods, editors, *Handbook of the History of Logic*, volume 10, Inductive Logic, pages 265–310. Elsevier/North-Holland, Amsterdam, 2010.