The Principle of Signature Exchangeability

Ronel, Tahel and Vencovska, Alena

2015

MIMS EPrint: **2015.76**

ISSN 1749-9097

The Principle of Signature Exchangeability

Tahel Ronel[1], Alena Vencovská[2,*]

School of Mathematics, The University of Manchester, Manchester M13 9PL

Abstract

We investigate the notion of a *signature* in Polyadic Inductive Logic and study the probability functions satisfying the *Principle of Signature Exchangeability*. In the binary case, we prove a representation theorem for such functions and show that they satisfy a binary version of the Principle of Instantial Relevance. We discuss polyadic versions of the Principle of Instantial Relevance and Johnson's Sufficientness Postulate.

Keywords: Constant Exchangeability, Pure Inductive Logic, Polyadic Atoms, Polyadic Signature, Instantial Relevance, Johnson's Sufficientness Postulate, Logical Probability, Rationality, Uncertain Reasoning.

1. Introduction

This paper is set in Pure Inductive Logic (PIL), see for example [10], which the reader is referred to for background and an extensive bibliography. In this subject, we are concerned with assigning subjective probabilities to sentences of a language according to rational considerations, traditionally based on the notions of symmetry, relevance and irrelevance.

The principle of Constant Exchangeability (Ex) or in Carnap's terms, the Axiom of Symmetry [1, 3], 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

*Corresponding author.

Email addresses: `tahel.ronel@manchester.ac.uk` (Tahel Ronel), `alena.vencovska@manchester.ac.uk` (Alena Vencovská)

[1]Supported by a UK Engineering and Physical Sciences Research Council PhD Grant.

[2]Supported by a UK Engineering and Physical Sciences Research Council Research Grant.

should be treated equally. In the usual framework of Inductive Logic 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 5) 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 (see [10, 12]), in this account we generalise the notion of a signature to polyadic Inductive Logic and investigate the theory this yields for higher arity languages. We begin by giving a brief account of the unary portion we shall be concerned with for the purpose of this paper, then suggest new methods and formulations for these concepts for general polyadic languages. Specifically, we present a polyadic 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 polyadic versions of the Principle of Instantial Relevance and Johnson's Sufficientness Postulate. We present this initially for languages with at most binary relation symbols and then, in the second part of the paper, we focus on the general case.

The context of this paper is as follows. We work with a first order language L containing finitely many relation symbols $R_1, \ldots, R_q$ of arities $r_1, \ldots, r_q$ respectively and countably many constant symbols $a_1, a_2, a_3, \ldots$, using the usual logical connectives and quantifiers. SL denotes the set of all sentences of the language L and $QFSL$ the set of all quantifier free sentences of the language. $b_1, \ldots, b_n$ or sometimes also $b'_1, \ldots, b'_n$ are used to denote some distinct constants from amongst the $a_1, a_2, \ldots$, and S_n stands for the set of permutations of $\{1, 2, \ldots, n\}$.

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. In addition, since we are interested in formulae only up to logical equivalence, we will often use '$=$' in place of '$\equiv$'.

Definition. A function $w : SL \to [0, 1]$ is a *probability function* if for all

θ, ϕ and $\exists x\, \psi(x) \in SL$
(P1) If θ is logically valid then $w(\theta) = 1$.
(P2) If θ and ϕ are mutually exclusive then $w(\theta \vee \phi) = w(\theta) + w(\phi)$.
(P3) $w(\exists x\, \psi(x)) = \lim_{n\to\infty} w(\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n))$.

Probability functions have a number of desirable properties, see for example [10]; note in particular that logically equivalent sentences always get the same probability.

The conditional probability of θ given ϕ, for ϕ such that $w(\phi) \neq 0$, is defined as follows:

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

We adopt the convention that expressions like $w(\theta \mid \phi) = a$ stand for $w(\theta \wedge \phi) = a\, w(\phi)$ even if $w(\phi) = 0$.

Any w satisfying just (P1) and (P2) on the quantifier free sentences of L has a unique extension to a probability function on SL, see [6], so in many situations it suffices to think of probability functions as defined on quantifier free sentences only, and satisfying (P1) and (P2).

As explained in [10, Chapter 7], this can be further reduced to a special class of such sentences called *state descriptions*, that is, to sentences $\Theta(b_1, \ldots, b_m)$ of the form

$$\bigwedge_{i=1}^{q} \quad \bigwedge_{\langle j_1 \ldots, j_{r_i}\rangle \in \{1,\ldots,m\}^{r_i}} \quad \pm\, R_i(b_{j_1} \ldots, b_{j_{r_i}}) \tag{1}$$

where $\pm R_i(b_{j_1} \ldots, b_{j_{r_i}})$ denotes one of $R_i(b_{j_1} \ldots, b_{j_{r_i}})$, $\neg R_i(b_{j_1} \ldots, b_{j_{r_i}})$. Furthermore, any w defined on state descriptions $\Theta(a_1, a_2, \ldots, a_m)$, $m \in \mathbb{N}$ to satisfy
(i) $w(\Theta(a_1, a_2, \ldots, a_m)) \geq 0$,
(ii) $w(\top) = 1$,
(iii) $w(\Theta(a_1, a_2, \ldots, a_m)) = \sum_{\Phi(a_1,a_2,\ldots,a_{m+1}) \models \Theta(a_1,a_2,\ldots,a_m)} w(\Phi(a_1, a_2, \ldots, a_{m+1}))$

extends uniquely to a probability function on $QFSL$ and hence on SL.

If $\Theta(b_1, \ldots, b_m)$ is a state description then $\Theta(x_1, \ldots, x_m)$ is called a *state formula*. We use the capital Greek letters Θ, Φ, Ψ for state descriptions and state formulae.

For a state description $\Theta(b_1, \ldots, b_m)$ and distinct $k_1, \ldots, k_g$ from $\{1, \ldots, m\}$,

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

or simply $\Theta[b_{k_1}, \ldots, b_{k_g}]$, denotes the restriction of $\Theta(b_1, \ldots, b_m)$ to $b_{k_1}, \ldots, b_{k_g}$. That is, the conjunction of the literals from (1) with $\{j_1, \ldots, j_{r_i}\} \subseteq \{k_1, \ldots, k_g\}$.

When the language is r-ary, the state formulae for r variables are called (polyadic) *atoms*, see [12] or [10]. In the case of a unary language, the atoms are the conjunctions $\bigwedge_{i=1}^{q} \pm R_i(x)$ and they are usually denoted $\alpha_1(x), \ldots, \alpha_{2^q}(x)$.

The idea of atoms has played an essential role in the study of Unary Inductive Logic since its conception by Johnson and Carnap, even if their formal expression of it differed [7, 2]. In particular, unary atoms have been used to formulate and investigate basic principles of the subject. This is possible, since unary state descriptions are the conjunctions of (instantiated) atoms,

$$\Theta(b_1, \ldots, b_m) = \bigwedge_{j=1}^{m} \alpha_{h_j}(b_j) \tag{2}$$

(where $h_j \in \{1, \ldots, 2^q\}$), and thus unary atoms form the basic building blocks of all sentences of a unary language.

Some Basic Principles of Unary Inductive Logic

The Principle of Constant Exchangeability is usually stated for a general (not necessarily unary) language as follows:

Constant Exchangeability, Ex *Let $\theta(a_1, \ldots, a_m) \in SL$ and let $b_1, \ldots, b_m$ be any other choice of distinct constant symbols from amongst the $a_1, a_2, \ldots$. Then*

$$w(\theta(a_1, \ldots, a_m)) = w(\theta(b_1, \ldots, b_m))\,. \tag{3}$$

It can be equivalently expressed as requiring (3) to hold only for state descriptions Θ instead of general $\theta \in SL$, see [10, Chapter 8]. This leads to a simpler formulation of Ex for unary languages (as mentioned above), based on the notion of a *signature*. The signature of a state description Θ as in (2) is defined to be the vector $\langle m_1, \ldots, m_{2^q}\rangle$ where m_i is the number of times that α_i appears amongst the α_{h_j}. Ex in the unary case thus amounts to

Constant Exchangeability, unary version *The probability of a state description depends only on its signature.*

We now mention a collection of important principles from Unary Inductive Logic that are stated in terms of (unary) atoms.

Atom Exchangeability, Ax *Let* $\Theta(b_1, \ldots, b_m) = \bigwedge_{j=1}^m \alpha_{h_j}(b_j)$ *be a state description and* $\sigma \in S_{2^q}$*. Then*

$$w\left(\bigwedge_{j=1}^m \alpha_{h_j}(b_j)\right) = w\left(\bigwedge_{j=1}^m \alpha_{\sigma(h_j)}(b_j)\right).$$

This principle can be equivalently expressed as requiring that state descriptions $\bigwedge_{j=1}^m \alpha_{h_j}(b_j)$ and $\bigwedge_{j=1}^m \alpha_{g_j}(b_j)$ satisfying

$$h_j = h_l \Longleftrightarrow g_j = g_l$$

must have the same probability.

Principle of Instantial Relevance, PIR

$$w\Bigg(\alpha_i(a_{m+2}) \mid \bigwedge_{j=1}^m \alpha_{h_j}(a_j)\Bigg) \leq w\Bigg(\alpha_i(a_{m+2}) \mid \alpha_i(a_{m+1}) \wedge \bigwedge_{j=1}^m \alpha_{h_j}(a_j)\Bigg).$$

This principle was suggested by Carnap [2, Chapter 13] and expresses the idea that having witnessed an event in the past should enhance (or at least should not decrease) our belief that we might see it again in future.

Johnson's Sufficientness Postulate, JSP $w\Big(\alpha_i(a_{m+1}) \mid \bigwedge_{j=1}^m \alpha_{h_j}(a_j)\Big)$ *depends only on* m *and on* m_i*, where* m_i *is the number of times that* α_i *appears amongst the* α_{h_j}*.*

First appearing in [7], 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.

Principle of Induction, PI *Assume that* $m_i \leq m_s$*, where* m_i*,* m_s *are the numbers of times that* α_i*,* α_s *respectively appear amongst the* α_{h_j}*. Then*

$$w\Bigg(\alpha_i(a_{m+1}) \mid \bigwedge_{j=1}^m \alpha_{h_j}(a_j)\Bigg) \leq w\Bigg(\alpha_s(a_{m+1}) \mid \bigwedge_{j=1}^m \alpha_{h_j}(a_j)\Bigg).$$

This principle [10, 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 [10, 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

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

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

$$w_{\vec{x}} \left(\bigwedge_{j=1}^{m} \alpha_{h_j}(b_j) \right) \;=\; \prod_{j=1}^{m} x_{h_j} \;=\; \prod_{i=1}^{2^q} x_i^{m_i},$$

where m_i is again the number of times that α_i appears amongst the α_{h_j}. Thus $w_{\vec{x}}$ is the (unique) function that assigns the probability x_i to all $\alpha_i(a_j)$ regardless of j, 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 [5] 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, [6]), and that - provided the language has at least two predicate symbols - any probability function satisfying Ex and JSP must be one of rather special functions called the Carnap Continuum functions ([7], and others). A later result due to Paris and Waterhouse [9] shows that any probability function satisfying Ex and Ax must also satisfy PI.

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 PI.

2. An Atom-based Approach for Binary Languages

We shall now consider how atoms can aid us to understand the properties of probability functions in the case when r is 2. That is, when L contains some binary relation symbols and possibly some unary predicate symbols, but no symbols of higher arity. We shall denote the unary predicate symbols by $P_1, \dots, P_{q_1}$ and the binary symbols by $Q_1, \dots, Q_{q_2}$ (rather than by R_i as we do for a general language), with $q_1 + q_2 = q$.

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

$$\bigwedge_{i=1}^{q_1} \pm P_i(x) \wedge \bigwedge_{i=1}^{q_2} \pm Q_i(x,x)$$

and we will write

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

for them (in some fixed order). We also refer to these formulae as *1-atoms.*

The *atoms* of the language[3], that is, the state formulae for two variables, have the form

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

[3] We sometimes say *binary atoms* for emphasis or, in some contexts later on, also 2-atoms. Note that 2-atoms will mean the same thing as atoms just when the language is binary.

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

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

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

$$\beta_k(x) \wedge \beta_c(y) \wedge \delta_e(x,y) \wedge \delta_d(y,x) \tag{4}$$

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

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

We refer to $\beta_k(x) \wedge \beta_c(y)$ as the *unary trace* of the atom (4).

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

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

and it is represented by the matrix

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

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

$$\Theta(b_1,\ldots,b_m) = \bigwedge_{1\leq i,t\leq m} \gamma_{h_{i,t}}(b_i,b_t) \tag{5}$$

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

$$\begin{pmatrix} k_1 & e_{1,2} & e_{1,3} & \ldots & e_{1,m} \\ d_{1,2} & k_2 & e_{2,3} & \ldots & e_{2,m} \\ d_{1,3} & d_{2,3} & k_3 & \ldots & e_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1,m} & d_{2,m} & d_{3,m} & \ldots & k_m \end{pmatrix} \tag{6}$$

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 $\Theta(b_1, \ldots, b_m) \models \delta_e(b_i, b_i)$.

Clearly, there is much over-specification in the expression (5); 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) to $i < t$,

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

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

$$\Theta(b_1, \ldots, b_m) = \bigwedge_{1 \le i \le m} \beta_{k_i}(b_i) \wedge \bigwedge_{1 \le i < t \le m} \gamma_{h_{i,t}}(b_i, b_t). \tag{7}$$

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. For Θ as in (7), we define

$$\bigwedge_{1 \le i \le m} \beta_{k_i}(b_i) \tag{8}$$

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*[4] rather than atoms (where elements in the binary case are the β_k

[4]See [11].

and δ_s, and analogously for higher arity languages), and thus to avoid over-specification. 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 although there are some advantages to doing this, some crucial connections are lost.

Using the alternative formulation of the unary Atom Exchangeability principle from page 5, and (5), it is straightforward to see how to formulate a binary counterpart of the Atom Exchangeability principle. The same approach works also for higher arity languages. This was investigated in [12] and it appears also in [10], so we will not pursue it in the present paper any further.

For the other principles 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. A *partial state description for* $b_1, \ldots, b_m$ *is a sentence*

$$\Delta(b_1, \ldots, b_m) = \bigwedge_{1 \leq i \leq m} \beta_{k_i}(b_i) \wedge \bigwedge_{\substack{\{b_i, b_t\} \in C \\ i<t}} \gamma_{h_{i,t}}(b_i, b_t)\,, \tag{9}$$

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

$$\begin{matrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{matrix}$$

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

$$\begin{array}{cccc} 1 & & 1 & \\ & 1 & & \\ 2 & & 1 & 2 \\ & & 2 & 1 \end{array}$$

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

$$\begin{array}{cccc} 1 & & 1 & \\ & 1 & 2 & \\ 2 & & 1 & 2 \\ & & 2 & 1 \end{array}$$

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 (9) contains no 2-element subsets, that is $C = \emptyset$, then (9) is still a partial state description. In particular, a unary trace for $b_1, \ldots, b_m$ is a partial state description for $b_1, \ldots, b_m$. Secondly, we mention that *partial state formulae* are defined analogously to partial state descriptions, with $b_1, \ldots, b_m$ replaced by (distinct) variables $x_1, \ldots, x_m$.

2.1. 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

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

This means that rather than N different ways a pair can behave, there are $p < N$ of them, where p is the number[5] of $\sim$-equivalence classes.

It will be convenient to introduce notation for these equivalence classes; we shall denote them by $\Gamma_1, \ldots, \Gamma_p$. From above, it follows that each class is

$$\left\{ \begin{pmatrix} k & e \\ d & c \end{pmatrix}, \begin{pmatrix} c & d \\ e & k \end{pmatrix} \right\} \tag{10}$$

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 (10) for some e, d.

Within the equivalence class (10), the unary trace of an atom determines the atom, except when $k = c$ and $e \neq d$. 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. For a state description

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

we define the *signature of* Θ to be the vector $\langle n_1, \ldots, n_p \rangle$, 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 (6) and Γ_j is (10), then n_j is the number of times one of the atoms from (10) appears as a submatrix of (6).

[5] Explicitly, $p = (N + 2^q \cdot 2^{q_2})/2$.

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

$$\sum_{k=1}^{2^q} m_k = m\,, \tag{11}$$

for $k \neq c$

$$\sum_{j \in A(k,c)} n_j = m_k m_c\,, \tag{12}$$

and

$$\sum_{j \in A(k,k)} n_j = \frac{m_k(m_k - 1)}{2}\,. \tag{13}$$

Conversely, thinking about state descriptions in terms of matrices as in (6), we can see that any $\vec{m}\vec{n} = \langle m_1, \dots, m_{2^q}\,;\, n_1, \dots, n_p \rangle$ such that (12) and (13) hold, is an extended signature of some $\Theta(b_1, \dots, b_m)$ for m defined by (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 clearly still implies Ex but the converse implication does not hold: 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.

The state descriptions represented by

$$\begin{array}{cccc} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{array} \qquad\qquad \begin{array}{cccc} 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 \\ 2 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 \end{array}$$

have the same signature but there are probability functions satisfying Ex that give these state descriptions different probabilities. For example, $u^{\bar{p},L}$ with $\bar{p} = \langle 0, \frac{1}{2}, \frac{1}{2}, 0, 0, \ldots \rangle$, see [10, Chapter 29], is one such function.

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 6) which satisfy BEx. These functions are characterised by an independence principle similar to the Constant Independence Principle (IP). In addition, 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, yields a proof of a binary generalisation of the Principle of Instantial Relevance, and a characterisation of a binary Carnap Continuum as the unique functions satisfying a binary generalisation of Johnson's Sufficientness Postulate. We begin with independence.

2.2. Binary Independence

The Constant Independence Principle IP (for any language), see page 6, requires that any two quantifier free sentences which have no constants 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. 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 a_i and a_j are brought together instantiating a relation in ψ.

That is, 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*

$$w(\phi \wedge \psi \,|\, \Delta) \;=\; w(\phi \,|\, \Delta) \cdot w(\psi \,|\, \Delta). \tag{14}$$

If $s = 0$ (the sentences have no constants in common) then $\Delta = \top$ (tautology), so BIP implies IP.

We shall now define the binary versions $w_{\vec{Y}}$ of the unary $w_{\vec{x}}$ mentioned on page 6. 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 $1 \leq k, c \leq 2^q$,

$$\sum_{j \in A(k,c)} s_j y_j = 1 \tag{15}$$

($A(k,c)$ was defined on page 12). 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

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

and Γ_j its equivalence class - that is, Γ_j is (10) - then

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

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, (15) says that the sum over z_h for those γ_h with a given unary trace is 1.

For a state description

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

we define

$$w_{\vec{Y}}(\Theta(a_1,\ldots,a_m)) = \prod_{1\le i\le m} x_{k_i} \prod_{1\le i<t\le m} z_{h_{i,t}}\,. \tag{16}$$

Note that if $\sigma \in S_m$ and

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

then $w_{\vec{Y}}(\Psi) = w_{\vec{Y}}(\Theta)$ since[6]

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

the multiset $\{k_{\sigma^{-1}(i)} : 1\le i\le m\}$ equals the multiset $\{k_i : 1\le i\le m\}$, and the multisets $\{h_{\sigma^{-1}(i),\sigma^{-1}(t)} : 1\le i<t\le m\}$, $\{h_{i,t} : 1\le i<t\le m\}$ 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 $\gamma_{h_{i',t'}} \sim \gamma_{h_{t',i'}}$ so $z_{h_{i',t'}} = z_{h_{t',i'}}$ and consequently $w_{\vec{Y}}(\Psi)$, $w_{\vec{Y}}(\Theta)$ must be equal.

Theorem 1. *Let L be a binary language. The functions $w_{\vec{Y}}$ defined above determine probability functions on SL that satisfy BEx and BIP (and hence also Ex and IP).*

Furthermore, any probability function satisfying Ex and BIP is equal to $w_{\vec{Y}}$ for some $\vec{Y}$.

Proof. $w = w_{\vec{Y}}$ clearly satisfies properties (i), (ii) from page 3. To show that it satisfies (iii), let $\Theta(a_1,\ldots,a_m)$ be as above. The $\Theta^+(a_1,\ldots,a_m,a_{m+1})$ which extend $\Theta(a_1,\ldots,a_m)$ have the form

$$\Theta(a_1,\ldots,a_m)\ \wedge\ \beta_c(a_{m+1})\ \wedge\ \bigwedge_{i=1}^{m} \gamma_{h_{i,m+1}}(a_i,a_{m+1})\,, \tag{17}$$

[6] Recall the convention from page 9 needed below when $\sigma^{-1}(i) > \sigma^{-1}(t)$.

where c is any of $1, \ldots, 2^q$ and the unary trace of $\gamma_{h_{i,m+1}}(x, y)$ is $\beta_{k_i}(x) \wedge \beta_c(y)$. The value of $w_{\vec{Y}}(\Theta^+)$ for Θ^+ as in (17) is

$$\left(\prod_{1 \leq i \leq m} x_{k_i} \right) x_c \left(\prod_{1 \leq i < t \leq m} z_{h_{i,t}} \right) \left(\prod_{i=1}^{m} z_{h_{i,m+1}} \right).$$

For a given c and for each $i = 1, \ldots, m$, the sum of the eligible $z_{h_{i,m+1}}$ is 1, since we are summing over all the $z_{h_{i,m+1}}$ such that $\gamma_{h_{i,m+1}}$ has trace $\beta_{k_i}(x) \wedge \beta_c(y)$. So summing the $w_{\vec{Y}}(\Theta^+)$ successively over these $h_{i,m+1}$ and then over c yields $w_{\vec{Y}}(\Theta)$. Hence (iii) holds, too. It follows that $w_{\vec{Y}}$ extends to a probability function on SL which moreover, by the remark preceding the theorem, satisfies Ex.

A similar argument now shows that (16) 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 $a_1, \ldots, a_M$). Thus $w_{\vec{Y}}$ satisfies BEx since the right hand side of (16) depends only on the signature of Θ.

To show BIP, we note that continuing with the same reasoning, we can show also that for a partial state description

$$\Phi(b_1, \ldots, b_m) = \bigwedge_{1 \leq i \leq m} \beta_{k_i}(b_i) \wedge \bigwedge_{\substack{\{b_i, b_t\} \in C \\ i < t}} \gamma_{h_{i,t}}(b_i, b_t) \tag{18}$$

we have

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

Assume that Φ and Ψ are some partial state descriptions which instantiate no pairs in common. Let $b_1, \ldots, b_s$ 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 we clearly have

$$w_{\vec{Y}}(\Phi \wedge \Psi \mid \Delta) = w_{\vec{Y}}(\Phi \mid \Delta)\, w_{\vec{Y}}(\Psi \mid \Delta) \tag{19}$$

because both sides are 0. So suppose Φ is as in (18), $s \leq m$,

$$\Psi(b_1, \ldots, b_s, b_{m+1}, \ldots, b_{m+n}) = \bigwedge_{1 \leq i \leq s} \beta_{k_i}(b_i) \wedge \bigwedge_{m+1 \leq i \leq m+n} \beta_{k_i}(b_i) \wedge \bigwedge_{\substack{\{b_i, b_t\} \in D \\ i<t}} \gamma_{h_{i,t}}(b_i, b_t) \quad (20)$$

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

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

We can now use the above observation regarding values of $w_{\vec{Y}}$ for partial state descriptions to prove that (19) holds in this case, too, since both sides are

$$\prod_{s+1 \leq i \leq m+n} x_{k_i} \prod_{\substack{\{b_i, b_t\} \in C \cup D \\ i<t}} z_{h_{i,t}}.$$

Hence BIP holds when ϕ, ψ are partial state descriptions.

To prove that (14) holds with general $\phi, \psi \in QFSL$, note that any quantifier free sentence $\phi(b_1, \ldots, b_m)$ is equivalent to a disjunction of partial state descriptions Φ_u as in (18), 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 (20) where $D = C^2_\psi$, and so by the above, for any unary trace Δ for $b_1, \ldots, b_s$,

$$w_{\vec{Y}}(\, \phi \wedge \psi \,|\, \Delta) \;=\; w_{\vec{Y}} \left(\bigvee_u \Phi_u \wedge \bigvee_f \Psi_f \,\middle|\, \Delta \right) \;=\; \sum_{u,f} w_{\vec{Y}}(\, \Phi_u \wedge \Psi_f \,|\, \Delta) =$$

$$\sum_{u,f} w_{\vec{Y}}\, (\Phi_u \mid \Delta) \cdot w_{\vec{Y}}\, (\Psi_f \mid \Delta) \;=\; \sum_u w_{\vec{Y}}\, (\Phi_u \mid \Delta) \cdot \sum_f w_{\vec{Y}}\, (\Psi_f \mid \Delta) =$$

$$w_{\vec{Y}} \left(\bigvee_u \Phi_u \,\middle|\, \Delta \right) \cdot w_{\vec{Y}} \left(\bigvee_f \Psi_f \,\middle|\, \Delta \right) = w_{\vec{Y}}(\phi \,|\, \Delta) \cdot w_{\vec{Y}}(\psi \,|\, \Delta),$$

as required.

For the final part of the theorem, assume that w satisfies Ex and BIP. We define

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

and

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

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 can 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. $\square$

2.3. Representation Theorem

We showed in Theorem 1 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 2. *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 $\theta \in SL$,*

$$w(\theta) = \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta) d\mu(\vec{Y}). \tag{21}$$

Conversely, for a given measure μ on the Borel subsets of $\mathbb{D}_L$, the function defined by (21) is a probability function on SL satisfying BEx.

Proof. Let w be a probability function for L satisfying BEx. It suffices to prove (21) for state descriptions, the rest follows, for instance, as in Corollary 9.2 of [10]. The proof is based on the fact that for a state description $\Theta(b_1, \ldots, b_m)$ and $u > m$

$$w(\Theta(b_1, \ldots, b_m)) = \sum_{\Psi(b_1,\ldots,b_m,b_{m+1},\ldots,b_u) \models \Theta(b_1,\ldots,b_m)} w(\Psi(b_1, \ldots, b_m, b_{m+1}, \ldots, b_u))\,, \tag{22}$$

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

$$\binom{t}{\{t_i : i \in \{1, \dots, n\}\}} = \binom{t}{t_1, t_2, \dots, t_n} = \frac{t!}{t_1! t_2! \dots t_n!}.$$

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. First, 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 (6), 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 $\vec{u}\vec{t}$, denoted by $\mathcal{N}(\emptyset, \vec{u}\vec{t})$, is

$$\binom{u}{u_1, \dots, u_{2^q}} \prod_{1 \leq k < c \leq 2^q} \binom{u_k u_c}{\{t_j : j \in A(k,c)\}} \times \prod_{1 \leq k \leq 2^q} \left(\binom{\frac{u_k(u_k-1)}{2}}{\{t_j : j \in A(k,k)\}} \prod_{j \in A(k,k)} s_j^{t_j} \right). \quad (23)$$

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

$$\binom{u-m}{u_1 - m_1, \dots, u_{2^q} - m_{2^q}} \prod_{1 \leq k < c \leq 2^q} \binom{u_k u_c - m_k m_c}{\{t_j - n_j : j \in A(k,c)\}} \times \prod_{1 \leq k \leq 2^q} \left(\binom{\frac{u_k(u_k-1)}{2} - \frac{m_k(m_k-1)}{2}}{\{t_j - n_j : j \in A(k,k)\}} \prod_{j \in A(k,k)} s_j^{(t_j - n_j)} \right). \quad (24)$$

We make the convention that our multinomial expression is 0 if any of the terms are negative. Note that the number calculated in (24) depends only on the signature $\vec{m}\vec{n}$ and not on the particular choice of $\Theta(b_1,\ldots,b_m)$.

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. From (22)

$$1 = w(\top) = \sum_{\vec{u}\vec{t}\in Sign(u)} \mathcal{N}(\emptyset,\vec{u}\vec{t})\, w(\vec{u}\vec{t}),$$

$$w(\vec{m}\vec{n}) = \sum_{\vec{u}\vec{t}\in Sign(u)} \mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})\ w(\vec{u}\vec{t}),$$

and hence

$$w(\vec{m}\vec{n}) = \sum_{\vec{u}\vec{t}\in Sign(u)} \frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})} \mathcal{N}(\emptyset,\vec{u}\vec{t})\ w(\vec{u}\vec{t}). \tag{25}$$

We shall show that

$$\left| \left(\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})}\right) - \left(\prod_{1\le k\le 2^q} \left(\frac{u_k}{u}\right)^{m_k} \prod_{1\le k<c\le 2^q} \left(\prod_{j\in A(k,c)} \left(\frac{t_j}{u_k u_c}\right)^{n_j}\right) \prod_{1\le k\le 2^q} \left(\prod_{j\in A(k,k)} \left(\frac{t_j s_j^{-1}}{\left(\frac{u_k(u_k-1)}{2}\right)}\right)^{n_j}\right)\right)\right| \tag{26}$$

is of the order $\mathcal{O}\left(\sqrt{u}^{-1}\right)$ (independently of $u_1,\ldots,u_{2^q},t_1,\ldots,t_p$). 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.

First, let $m_k \le u_k$ and $n_j \le t_j$ for every j,k, so that none of the terms in (24) 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 (26) can be written as

$$\left(\prod_{1\le k\le 2^q} \left(\frac{u_k}{u}\right)^{m_k} \prod_{1\le k<c\le 2^q} \left(\prod_{j\in A(k,c)} \left(\frac{t_j}{u_k u_c}\right)^{n_j}\right) \prod_{1\le k\le 2^q} \left(\prod_{j\in A(k,k)} \left(\frac{t_j s_j^{-1}}{\left(\frac{u_k(u_k-1)}{2}\right)}\right)^{n_j}\right)\right) \tag{27}$$

$$\times \frac{\prod_{1\le k\le 2^q} \prod_{0\le i\le m_k-1} \left(1 - i\, u_k^{-1}\right)}{\prod_{0\le l\le m-1} \left(1 - l\, u^{-1}\right)} \tag{28}$$

$$\times \prod_{1\le k<c\le 2^q} \left(\frac{\prod_{j\in A(k,c)} \prod_{0\le i\le n_j-1} \left(1 - i\, t_j^{-1}\right)}{\prod_{0\le l\le m_k m_c-1} \left(1 - l\, \left(u_k u_c\right)^{-1}\right)} \right) \tag{29}$$

$$\times \prod_{1\le k\le 2^q} \left(\frac{\prod_{j\in A(k,k)} \prod_{0\le i\le n_j-1} \left(1 - i\, t_j^{-1}\right)}{\prod_{0\le l\le (m_k(m_k-1)/2)-1} \left(1 - l\, \left(\frac{u_k(u_k-1)}{2}\right)^{-1}\right)} \right). \tag{30}$$

Let P stand for the product of (28), (29) and (30).

We observe that P is bounded by a constant independent of u, the u_k and the t_j. For example,

$$(28) < \left(\frac{1}{1-(m-1)\, m^{-1}}\right)^m$$

and similarly for (29) and (30).

Furthermore, we need only consider those k where $m_k > 0$ in the limit of (26) 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.

We shall prove the claim about (26) by cases. Consider first the case that for some k with $m_k > 0$ we have $u_k \le \sqrt{u}$. Then

$$\prod_{1\le k\le 2^q} \left(\frac{u_k}{u}\right)^{m_k} \le (\sqrt{u})^{-1},$$

each of the other products in (27) is at most 1, so $(26) = |\,(27)\cdot(1-P)\,| = \mathcal{O}(\sqrt{u}^{-1})$. 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 $t_j \le \sqrt{u}$.

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 (28) can be written as a product of m fractions of the form $\frac{1-\alpha u_k{}^{-1}}{1-\beta u^{-1}}$, $\alpha, \beta \in \{1,\ldots,m\}$ and that the distance of each fraction from 1 is

$$\left|\frac{1-\alpha u_k{}^{-1}}{1-\beta u^{-1}} - 1\right| < 2\left(\beta u^{-1} + \alpha u_k{}^{-1}\right) < \; 2\sqrt{u}^{-1}\,(\alpha+\beta) \;\le\; 4m\sqrt{u}^{-1}.$$

Hence (28) is $1 + \mathcal{O}\left(\sqrt{u}^{-1}\right)$. A similar argument works for the other two products, (29) and (30), so P is $1 + \mathcal{O}\left(\sqrt{u}^{-1}\right)$. It follows that (26) 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 $\frac{\mathcal{N}(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}(\emptyset,\vec{u}\vec{t})} = 0$. In addition, $m_k > 0$ and $u_k < \sqrt{u}$, so arguing as above (27) would be of order $\mathcal{O}(\sqrt{u}^{-1})$ and consequently so would (26), which exhausts all cases.

Define $\vec{Y}_{\vec{u}\vec{t}}$ by

$$x_k = \frac{u_k}{u}, \qquad y_j = \begin{cases} \frac{t_j}{u_k u_c} & \text{for } j \in A(k,c),\ u_k, u_c \neq 0,\ k < c, \\ \frac{t_j {s_j}^{-1}}{\binom{\frac{u_k(u_k-1)}{2}}{}} & \text{for } j \in A(k,k),\ u_k \neq 0, 1, \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

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 (27).

We shall now employ methods from Nonstandard Analysis, particularly Loeb Measure Theory [8, 4] to complete the proof. An alternative classical proof may be found in [11].

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 (25)) we still have

$$w^*(\vec{m}\vec{n}) = \sum_{\vec{u}\vec{t} \in Sign^*(u)} \frac{\mathcal{N}^*(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}^*(\emptyset,\vec{u}\vec{t})} \mathcal{N}^*(\emptyset,\vec{u}\vec{t}) w^*(\vec{u}\vec{t}). \tag{32}$$

Loeb Measure Theory enables us to conclude from (32) that for some σ-additive measure μ' on $Sign^*(u)$ we have (for all standard extended signatures $\vec{m}\vec{n}$)

$$w(\vec{m}\vec{n}) = \int_{Sign^*(u)} {}^{\circ}\left(\frac{\mathcal{N}^*(\vec{m}\vec{n},\vec{u}\vec{t})}{\mathcal{N}^*(\emptyset,\vec{u}\vec{t})}\right) d\mu'(\vec{u}\vec{t})\,. \tag{33}$$

Since, in U, (26) is $\mathcal{O}\left(\sqrt{u}^{-1}\right)$, this gives

$$w(\vec{m}\vec{n}) = \int_{Sign^*(u)} {}^\circ(w^*_{\vec{u}\vec{t}}(\vec{m}\vec{n}))\; d\mu'(\vec{u}\vec{t})\,. \tag{34}$$

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 (31), means (34) becomes (using, for example, Proposition 1, Chapter 15 of [13])

$$w(\vec{m}\vec{n}) = \int_{\mathbb{D}_L} w_{\vec{Y}}(\vec{m}\vec{n})\, d\mu(\vec{Y}),$$

as required.

In the opposite direction, a function on SL defined by (21) 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. $\square$

We shall now use the above representation theorem to show that the $w_{\vec{Y}}$ functions, which by Theorem 1 are the only probability functions satisfying BIP and Ex, can be characterised alternatively as the only probability functions satisfying IP and BEx. The fact that the $w_{\vec{Y}}$ satisfy BEx and IP follows from Theorem 1 and the other part follows from the following theorem.

Theorem 3. *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. [7] Let μ be the σ-additive normalised measure guaranteed to exist by Theorem 2 such that

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

[7] We use the method of the proof of [10, Theorem 20.6].

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

$$\begin{aligned}
0 &= 2\,(w(\theta \wedge \theta') - w(\theta) \cdot w(\theta')) \\
&= \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta \wedge \theta')\, \mathrm{d}\mu(\vec{Y}) + \int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta \wedge \theta')\, \mathrm{d}\mu(\vec{Y'}) \\
&\quad -2\left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\, \mathrm{d}\mu(\vec{Y})\right) \cdot \left(\int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta')\, \mathrm{d}\mu(\vec{Y'})\right) \\
&= \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta) \cdot w_{\vec{Y}}(\theta')\, \mathrm{d}\mu(\vec{Y}) + \int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta) \cdot w_{\vec{Y'}}(\theta')\, \mathrm{d}\mu(\vec{Y'}) \\
&\quad -2\left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\, \mathrm{d}\mu(\vec{Y})\right) \cdot \left(\int_{\mathbb{D}_L} w_{\vec{Y'}}(\theta')\, \mathrm{d}\mu(\vec{Y'})\right) \\
&= \int_{\mathbb{D}_L} \int_{\mathbb{D}_L} \left(w_{\vec{Y}}(\theta) - w_{\vec{Y'}}(\theta')\right)^2\, \mathrm{d}\mu(\vec{Y})\, \mathrm{d}\mu(\vec{Y'}).
\end{aligned}$$

It follows that there exists a subset A of $\mathbb{D}_L$ with μ measure 1 such that $w_{\vec{Y}}(\theta)$ as a function of $\vec{Y}$ is constant on A for *every* $\theta \in QFSL$, see e.g. [11] for details. Therefore, for any $\vec{Y} \in A$ we must have that w and $w_{\vec{Y}}$ are equal for quantifier free sentences and hence for all sentences, as required. □

2.4. 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,

$$w(\beta_k(a_{m+2}) \,|\, \Delta) \;\leq\; w(\beta_k(a_{m+2}) \,|\, \beta_k(a_{m+1}) \wedge \Delta). \tag{35}$$

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. Clearly, 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 $\Delta(a_1, \ldots, a_m)$ be a partial state description. Then (35) 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*

$$w(\gamma_h(b_1, b_2) \mid \Delta) \leq w(\gamma_h(b_1, b_2) \mid \Delta \wedge \gamma_h(b_1', b_2'))\,. \tag{36}$$

Theorem 4. *Let w be a probability function on SL satisfying BEx. Then w satisfies the Binary Principle of Instantial Relevance.*

Proof. First, note that every $w_{\vec{Y}}$ satisfies (35) and (36) with equality by the definition of these functions. This is the case since

$$w_{\vec{Y}}(\Delta \wedge \beta_k(a_{m+2})) = w_{\vec{Y}}(\Delta) \cdot x_k\,,$$

$$w_{\vec{Y}}(\Delta \wedge \beta_k(a_{m+2}) \wedge \beta_k(a_{m+1})) = w_{\vec{Y}}(\Delta) \cdot x_k^2\,,$$

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

$$w(\Delta \wedge \gamma_h(b_1, b_2)) = w(\Delta \wedge \gamma_h(b_1', b_2')) = w(\Delta) \cdot y_h \cdot x_c\,,$$

$$w(\Delta \wedge \gamma_h(b_1, b_2) \wedge \gamma(b_1', b_2')) = w(\Delta) \cdot y_h^2 \cdot x_c^2 \,.$$

By Theorem 2, since w satisfies BEx, w is an integral of the $w_{\vec{Y}}$. Let μ be the corresponding measure. Then (35) and any instance of (36) become

$$\left(\int_{\mathbb{D}_L} f(\vec{Y}) w_{\vec{Y}}(\Delta) d\mu(\vec{Y})\right)^2 \leq \left(\int_{\mathbb{D}_L} (f(\vec{Y}))^2 w_{\vec{Y}}(\Delta) d\mu(\vec{Y})\right) \left(\int_{\mathbb{D}_L} w_{\vec{Y}}(\Delta) d\mu(\vec{Y})\right)$$

for some function f (in the above cases, $f(\vec{Y})$ is x_k or $y_h x_c$ respectively), and this integral inequality holds for any f, as required. $\square$

We remark that the same method yields the following related result:

Theorem 5. *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*

$$w(\gamma_h(b_1', b_2') \,|\, \Delta) \leq w(\gamma_h(b_1, b_2) \,|\, \Delta \wedge \gamma_h(b_1', b_2')) \,. \tag{37}$$

2.5. Binary Sufficientness Postulate

Finally, we turn our attention to disregarding irrelevant information. A classical principle in the unary case is Johnson's Sufficientness Postulate (see Page 5). This principle says that the conditional probability of a new constant satisfying an atom, given a state description, should not depend on what the other atoms in the state description do, only on how many constants the state description is about and how many of them satisfy this very same atom.

An earlier attempt at generalising Johnson's Sufficientness Postulate to the binary context was made in [14], but it proved very restrictive in the sense that only two probability functions satisfied it. This earlier approach was not based on atoms, but rather focused on the conditional probability of a full state description for $a_1, \ldots, a_{m+1}$ given a full state description for $a_1, \ldots, a_m$. Such an approach corresponds to assuming that an agent learns about the world through successively encountering new individuals and learning everything about each of them - including all their connections to all individuals encountered previously - in one go.

However, suppose instead that the agent only learns all about (one or) two individuals at a time and learns nothing about their connections to other individuals. That is, the agent is always focusing on (at most) two individuals at any one time. The fact that the language is binary makes this a plausible assumption. Then it is clear that the evidence should be a partial state description, and that we need to consider the conditional probability of an instantiated atom - some $\gamma_h(a_u, a_v)$, or a 1-atom - some $\beta_k(a_u)$.

Furthermore, for a partial state description

$$\Delta(a_1, \ldots, a_m) = \bigwedge_{1 \leq i \leq m} \beta_{k_i}(a_i) \wedge \bigwedge_{\substack{\{a_i, a_t\} \in C \\ i<t}} \gamma_{h_{i,t}}(a_i, a_t)\,,$$

the probability of an extension of it by some $\beta_k(a_{m+1})$ (how a new individual behaves in isolation), should arguably depend only on the β_{k_i} (how other individuals behave in isolation), rather than on the $\gamma_{h_{i,t}}$. An extension of Δ by some $\gamma_h(a_u, a_v)$ for $1 \leq u < v \leq m$, $\{a_u, a_v\} \notin C$ (how a_u and a_v relate to each other given how each of them behaves in isolation), should depend only on those $\gamma_{h_{i,t}}$ where a_i and a_t behave in isolation just as a_u and a_v do.

Accordingly, we broaden the notion of the extended signature of a state description to partial state descriptions. We define the *extended signature of* Δ as above, to be the vector $\langle m_1, \ldots, m_{2^q}\,;\, n_1, \ldots, n_p \rangle$, where m_k is the number of times that k appears amongst the k_i and n_j is the number of $\langle i, t\rangle$ such that $\{a_i, a_t\} \in C$ and $\gamma_{h_{i,t}} \in \Gamma_j$.

Recall that $j(h)$ denotes the j for which $\gamma_h \in \Gamma_j$ and $A(k, c)$ denotes the set of all $j = j(h)$ such that the unary trace of $\gamma_h(x, y)$ is $\beta_k(x) \wedge \beta_c(y)$. The extended signature of Δ still satisfies

$$\sum_{k=1}^{2^q} m_k = m\,, \tag{38}$$

but the sums of n_j for $j \in A(k, c)$ or $A(k, k)$ no longer need to be as in (12), (13) and we define

$$\sum_{j \in A(k,c)} n_j \;=\; n_{k,c} \qquad\qquad \sum_{j \in A(k,k)} n_j \;=\; n_{k,k}\,.$$

Note that $n_{k,c} = n_{c,k}$.

Taking our argument further along the lines of the Unary Johnson's Sufficientness Postulate, we are led to the requirement that

• *the conditional probability of $\beta_k(a_{m+1})$ given Δ should depend only on m_k and m,*

and noting that for $j \in A(k_u, k_v)$ there are s_j (that is, 1 or 2) atoms γ_h in Γ_j such that $\gamma_h(a_u, a_v)$ is consistent with Δ (because the unary trace is fixed), we further require that

• *for $1 \leq u < v \leq m$, $j \in A(k_u, k_v)$ and $h \in \Gamma_j$, the conditional probability of $\gamma_h(a_u, a_v)$ given Δ should depend only on n_j, s_j and n_{k_u,k_v}.*

The (atom-based) **Binary Sufficientness Postulate** consists of the two requirements above. It is shown in [15] that the unique regular[8] probability functions w satisfying Ex and this principle, are those for which the above conditional probabilities satisfy

$$w(\beta_k(a_{m+1}) \,|\, \Delta) = \frac{m_k + \frac{\mu}{2^q}}{m + \mu}\,, \qquad w(\gamma_h(a_u, a_v)|\Delta) = \frac{\frac{n_j}{s_j} + \frac{\lambda}{2^{2q_2}}}{n_{k_u,k_v} + \lambda}$$

for some $0 < \mu, \lambda \leq \infty$.

3. An Atom-based Approach for Polyadic Languages

For the rest of this paper, we assume again that L is a language with relation symbols $R_1, \dots, R_q$ of arities $r_1, \dots, r_q$. Moreover, we assume that it is r-ary for some $r > 1$, so the maximum of the r_i is r.

The atoms of L are the state formulae for r variables. We denote them[9]

$$\gamma_1(x_1, \dots, x_r), \dots, \gamma_N(x_1, \dots, x_r)\,.$$

As in the binary case, state descriptions for at least r constants can be expressed as a conjunction of (instantiated) atoms,

$$\Theta(b_1, \dots, b_m) = \bigwedge_{1 < i_1 < \ldots < i_r \leq m} \gamma_{h_{i_1,\ldots,i_r}}(b_{i_1}, \dots, b_{i_r})\,. \tag{39}$$

[8] Regular probability functions are those that give all state descriptions non-zero probabilities.

[9] Note that N (as well as the N_g defined below) depend on L.

Clearly, we have $\gamma_{h_{i_1,\dots,i_r}}(b_{i_1},\dots,b_{i_r}) = \Theta[b_{i_1},\dots,b_{i_r}]$; these sentences express all the information contained in Θ that involves $b_{i_1},\dots,b_{i_r}$ and no other constants. Polyadic atoms thus again act as the basic building blocks for all sentences of the language. Note that no 'smaller units' involving fewer variables could play this role, since the language is r-ary.

Even so, we will find it convenient to have a way of referring to blocks smaller than atoms.

Definition. The *g-atoms* for $g \leq r$ are the state formulae of L for g variables. They are denoted by

$$\gamma_1^g(x_1,\dots,x_g),\dots,\gamma_{N_g}^g(x_1,\dots,x_g)\,.$$

Thus the $\gamma_h^r(x_1,\dots,x_r)$ are just the atoms $\gamma_h(x_1,\dots,x_r)$ and $N_r = N$. 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 section to avoid superscripts altogether.

Every state description for at least r constants can be expressed as a conjunction (39). Conversely, such a conjunction is consistent (and hence defines a state description) just when any pair of the $\gamma_{h_{i_1,\dots,i_r}}(b_{i_1},\dots,b_{i_r})$ agree when restricted to the constants they have in common. We will find it useful to make these shared components visible so we write

$$\Theta(b_1,\dots,b_m) \;=\; \bigwedge_{1\leq s\leq r}\;\bigwedge_{1\leq i_1<\ldots<i_s\leq m}\gamma^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s})\,. \qquad (40)$$

This works even when $m < r$. Note that the $\gamma^s_{h_{i_1,\dots,i_s}}$ are such that

$$\gamma^s_{h_{i_1,\dots,i_s}}(b_{i_1},\dots,b_{i_s}) = \Theta[b_{i_1},\dots,b_{i_s}].$$

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 to the *g-ary trace* of the state description. More precisely,

Definition. The *g-ary trace* of the state description (40), denoted by

$$(\Theta{\restriction}g)(b_1,\dots,b_m),$$

or sometimes simply $(\Theta \restriction g)$, is defined to be

$$\bigwedge_{1\le s\le g} \bigwedge_{1\le i_1<\ldots<i_s\le m} \gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1},\ldots,b_{i_s})\,. \tag{41}$$

Note that this agrees with the definition of the unary trace we made in the previous section (on page 9). Any consistent conjunction of the form (41) 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. A *partial state description* for $b_1,\ldots,b_m$ is a sentence of the form

$$\Delta(b_1,\ldots,b_m) \;=\; \bigwedge_{1\le s\le r} \bigwedge_{\substack{\{b_{i_1},\ldots,b_{i_s}\}\in C^s\\ i_1<\ldots<i_s}} \gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1},\ldots,b_{i_s})\,, \tag{42}$$

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

We will assume that (42), like (40), 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 $i=1,\ldots,m$. We remark also that a partial state description (42) 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.

3.1. 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 $\gamma^g_h \sim \gamma^g_k$ if there exists a permutation $\sigma\in S_g$ such that

$$\gamma^g_h(x_1,\ldots,x_g) \equiv \gamma^g_k(x_{\sigma(1)},\ldots,x_{\sigma(g)}) \tag{43}$$

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 (43) 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 section, 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 (given g and j, these subclasses of Γ_j^g all have the same number of elements). In the binary case, we wrote just s_j for s_j^2. Thus 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. Furthermore, we define s^g to be the *total* number of g-atoms with a given $(g-1)$-ary trace. Note that this is independent of the trace chosen, and that for any given $(g-1)$-ary trace, s^g is the sum of the s_j^g over the j for which Γ_j^g contains an atom with this trace.

We extend the definition of a *signature* from binary languages to r-ary languages for $r > 2$ in the obvious way:

Definition. The *signature* of a state description Θ as in (39) (or (40)) 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 (40) 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 $\gamma_{h_{i_1,\ldots,i_g}}^g \in \Gamma_j^g$. 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 L unary or binary is the same as Ex or BEx respectively. Sgx implies Ex but the converse implication does not hold in general. We gave an example of a probability function satisfying Ex but not Sgx (BEx) for $r = 2$ in the previous section.

3.2. Polyadic Independence

The following definition aims to capture exactly which sets of g constants are brought together instantiating a relation within a sentence:

Definition. For a sentence $\phi(b_1, \dots, b_m) \in QFSL$ 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.

A modification of the Disjunctive Normal Form Theorem yields the following lemma:

Lemma 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* (42)*, 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 based on atoms, as stated on Page 6 for $r = 1$ and on page 15 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-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*

$$C^{g+1}_\phi \cap C^{g+1}_\psi = \emptyset$$

and let $b_1, \dots, b_t$ be the constants that ϕ and ψ have in common (if any). Let Δ be a g-trace for the constants $b_1, \dots, b_t$ when $t > 0$, and $\Delta = \top$ (tautology) if ϕ and ψ have no constants in common. Then

$$w(\phi \wedge \psi \,|\, \Delta) = w(\phi \,|\, \Delta) \cdot w(\psi \,|\, \Delta)\,.$$

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 $\vec{Y} = \langle y^1_1, \ldots, y^1_{p_1}\,;\, y^2_1, \ldots, y^2_{p_2}\,;\, \ldots\,;\, y^r_1, \ldots, y^r_{p_r}\rangle$ be a vector of real numbers such that

$$0 \leq y^g_j \leq 1, \qquad \sum_{j=1}^{p_1} y^1_j = 1,$$

and such that for $1 < g \leq r$ the following holds: For any $(g-1)$-ary trace ψ for $x_1, \ldots, x_g$,

$$\sum_j s^g_j\, y^g_j = 1 \tag{44}$$

where the sum is taken over those $j \in \{1, \ldots, p_g\}$ for which Γ^g_j contains some γ^g_h 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

$$z^g_h = y^g_{j(h)}$$

where $j(h)$ is that j for which $\gamma^g_h \in \Gamma^g_j$. Note that (44) is the same as requiring

$$\sum_{(\gamma^g_h \restriction g-1)=\psi} z^g_h \;=\; 1\,. \tag{45}$$

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^1_h are chosen for $b_1, \ldots, b_m$, independently according to the probabilities z^1_h. Then the γ^2_h 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^2_h for which $(\gamma^2_h \restriction 1)(x_1, x_2) \equiv \gamma^1_{h_{i_1}}(x_1) \wedge \gamma^1_{h_{i_2}}(x_2)$, independently and according to the probabilities z^2_h, and so on. Note that this works by virtue of (45), because when choosing γ^g_h for $b_{i_1}, \ldots, b_{i_g}$, $(\gamma^g_h \restriction g-1)$ is determined.

More formally, given $\vec{Y}$ as above, for a state description $\Theta(a_1, \ldots, a_m)$ such that

$$\Theta(a_1, \ldots, a_m) \equiv \bigwedge_{\substack{1 \leq s \leq r \\ 1 \leq i_1 < \ldots < i_s \leq m}} \gamma^s_{h_{i_1, \ldots, i_s}}(a_{i_1}, \ldots, a_{i_s}) \tag{46}$$

we define

$$w_{\vec{Y}}(\Theta(a_1,\ldots,a_m)) = \prod_{\substack{1\le s\le r\\ 1\le i_1<\ldots<i_s\le m}} z^s_{h_{i_1,\ldots,i_s}}\,. \tag{47}$$

Note that, as in the binary case, if $\sigma \in S_m$ and $\Psi(a_1,\ldots,a_m) = \Theta(a_{\sigma(1)},\ldots,a_{\sigma(m)})$ then $w_{\vec{Y}}(\Theta) = w_{\vec{Y}}(\Psi)$.

Theorem 7. *The functions $w_{\vec{Y}}$ determine probability functions that satisfy Sgx and SIP (and hence also Ex and IP).*

Furthermore, any probability function satisfying Ex and SIP is equal to $w_{\vec{Y}}$ for some $\vec{Y}$.

Proof. To show that $w_{\vec{Y}}$ determines a probability function note that (i) and (ii) from page 3 clearly hold. For (iii), 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 $\Theta^+(a_1,\ldots,a_m,a_{m+1})$ extend Θ. Then $w_{\vec{Y}}(\Theta^+(a_1,\ldots,a_m,a_{m+1}))$ is the product

$$\left(\prod_{\substack{1\le s\le r\\ 1\le i_1<\ldots<i_s\le m}} z^s_{h_{i_1,\ldots,i_s}}\right)\left(\prod_{\substack{1\le s\le r\\ 1\le i_1<\ldots<i_{s-1}\le m}} z^s_{h_{i_1,\ldots,i_{s-1},(m+1)}}\right)$$

where the first product is as for Θ and $h_{i_1,\ldots,i_{s-1},(m+1)}$ is that h for which

$$\gamma^s_h(a_{i_1},\ldots,a_{i_{s-1}},a_{m+1}) = \Theta^+[a_{i_1},\ldots,a_{i_{s-1}},a_{m+1}]\,.$$

That is, where Θ^+ is

$$\Theta(a_1,\ldots,a_m) \quad\wedge \bigwedge_{\substack{1\le s\le r\\ 1\le i_1<\ldots<i_{s-1}\le m}} \gamma^s_{h_{i_1,\ldots,i_{s-1},(m+1)}}(a_{i_1},\ldots,a_{i_{s-1}},a_{m+1}). \tag{48}$$

Consider some r-tuple $\langle i_1,\ldots,i_{r-1},(m+1)\rangle$ with $1\le i_1<\ldots<i_{r-1}\le m$. If some $\Theta^+\models\Theta$ satisfies

$$\Theta^+[a_{i_1},\ldots,a_{i_{r-1}},a_{m+1}] = \gamma^r_h(a_{i_1},\ldots,a_{i_{r-1}},a_{m+1})\,,$$

then any conjunction that differs from (48) only by having $\gamma_k^r(a_{i_1}, \ldots, a_{i_{r-1}}, a_{m+1})$ in place of $\gamma_h^r(a_{i_1}, \ldots, a_{i_{r-1}}, a_{m+1})$, where γ_h^r and γ_k^r have the same $(r-1)$-ary trace, is also a state description extending Θ. Since the z_k^r for all such k sum to 1 (from (45)), 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.

Similar reasoning gives us that (47) holds even when $a_1, \ldots, a_m$ are replaced by any other distinct constants $b_1, \ldots, b_m$, and that we have an analogous formula for the probability of partial state descriptions:

$$w_{\vec{Y}}\left(\bigwedge_{1\leq s\leq r}\ \bigwedge_{\substack{\{i_1,\ldots,i_s\}\in C^s\\ i_1<\ldots<i_s}} \gamma^s_{h_{i_1,\ldots,i_s}}(b_{i_1}, \ldots, b_{i_s})\right) = \prod_{1\leq s\leq r}\ \prod_{\substack{\{i_1,\ldots,i_s\}\in C^s\\ i_1<\ldots<i_s}} z^s_{h_{i_1,\ldots,i_s}}.$$

Using this, SIP is first seen to hold for partial state descriptions ϕ, ψ, and then employing Lemma 6, in general. We omit the details.

To prove the last part of the theorem, assume that w satisfies Ex and SIP. We define $\vec{Y}$ by

$$y^g_{j(h)} = z^g_h = w(\gamma^g_h(a_1, \ldots, a_m) \mid (\gamma^g_h(a_1, \ldots, a_m) \restriction g-1))$$

where $\gamma^g_h \in \Gamma^g_j$ and $(\gamma^g_h(a_1, \ldots, a_m) \restriction g-1)$ 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 (44) must hold. Writing any state description in the form (46) and using Ex and SIP, we can show by induction (adding the conjuncts for increasing numbers of constants one by one) that its probability is given by (47). □

Corollary 8. *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*

$$w(\theta) \ = \ \int_{\mathbb{D}_L} w_{\vec{Y}}(\theta)\, d\mu(\vec{Y}). \tag{49}$$

Then the function w is a probability function on SL satisfying Sgx.

Proof. This can be proved by a straightforward checking of (P1), (P2), (P3) and Sgx, using the Lebesgue Dominated Convergence Theorem for (P3). □

However, whether or not the converse to Corollary 8 holds, that is, whether any probability function satisfying Sgx can be expressed in the form (49) remains to be investigated.

3.3. Polyadic PIR and JSP

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

$$\Delta(a_1, \ldots, a_m) \;=\; \bigwedge_{1 \le s \le r} \;\; \bigwedge_{\substack{\{a_{i_1}, \ldots, a_{i_s}\} \in C^s \\ i_1 < \ldots < i_s}} \gamma^s_{h_{i_1, \ldots, i_s}}(a_{i_1}, \ldots, a_{i_s}) \qquad (50)$$

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. 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 $\{b_{i_1}, \ldots, b_{i_{r_d}}\} \in E$.

Polyadic Principle of Instantial Relevance, PPIR *Let* $\Delta(a_1, \ldots, a_m)$ *be a partial state description,* $1 \le g \le r$*, 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*

$$w(\gamma^g_h(b_1, \ldots, b_g) \,|\, \Delta) \le w(\gamma^g_h(b_1, \ldots, b_g) \,|\, \Delta \wedge \gamma^g_h(b'_1, \ldots, b'_g))\,. \qquad (51)$$

Theorem 9. *Any convex combination (or integral) of the functions* $w_{\vec{Y}}$ *satisfies PPIR.*

Proof. Let Δ, γ^g_h and $b_1, \ldots, b_g$, $b'_1, \ldots, b'_g$ be as in the statement of PPIR. Assume Δ is as in (50). Let E be the Extra in $\gamma^g_h(b_1, \ldots, b_g)$ over Δ. We have

$$w_{\vec{Y}}(\Delta) = \prod_{1\le s\le r} \prod_{\substack{\{a_{i_1},\ldots,a_{i_s}\}\in C^s \\ i_1<\ldots<i_s}} z^s_{h_{i_1,\ldots,i_s}} ,$$

$$w_{\vec{Y}}(\Delta \wedge \gamma^g_h(b_1,\ldots,b_g)) = w_{\vec{Y}}(\Delta \wedge \gamma^g_h(b'_1,\ldots,b'_g)) = w_{\vec{Y}}(\Delta)\cdot \prod_{\{t_1,\ldots,t_s\}\in E} z^s_{k_{t_1,\ldots,t_s}} ,$$

$$w_{\vec{Y}}(\Delta \wedge \gamma^g_h(b_1,\ldots,b_g) \wedge \gamma^g_h(b'_1,\ldots,b'_g)) = w_{\vec{Y}}(\Delta)\cdot \left(\prod_{\{t_1,\ldots,t_s\}\in E} z^s_{k_{t_1,\ldots,t_s}}\right)^2 .$$

It follows that for $w = w_{\vec{Y}}$, (51) holds with equality.

The proof for w defined by (49), and hence also for any convex combination of the $w_{\vec{Y}}$, follows from the above equations exactly as in the binary case. □

By the same method we also obtain that under the same assumptions as those in PPIR except that merely the Extras in $\gamma^g_h(b_1,\ldots,b_g)$ over $\Delta\wedge\gamma^g_h(b'_1,\ldots,b'_g)$ and in $\gamma^g_h(b'_1,\ldots,b'_g)$ over Δ are required to be the same, we obtain that any convex combination (or integral) w of the functions $w_{\vec{Y}}$ satisfies

$$w(\gamma^g_h(b'_1,\ldots,b'_g)\,|\,\Delta) \le w(\gamma^g_h(b_1,\ldots,b_g)\,|\,\Delta\wedge\gamma^g_h(b'_1,\ldots,b'_g)) .$$

We now address irrelevance. Searching for a polyadic variant of JSP, we are again led to consider the conditional probability of some $\gamma^g_h(b_1,\ldots,b_g)$ given a partial state description $\Delta(a_1,\ldots,a_m)$ as in (50). For $1 < g \le r$, we will require the partial state description to be $(g-1)$-*complete*, that is, C^{g-1} contain all the $(g-1)$-element subsets of $\{a_1,\ldots,a_m\}$. In other words, Δ implies a state description for any $(g-1)$-tuple of constants from amongst the $a_1,\ldots,a_m$.[10] Furthermore, we will require that the Extra in $\gamma^g_h(b_1,\ldots,b_g)$ over Δ contains just the set $\{1,2,\ldots,g\}$, that is, either $g=1$ and we consider the conditional probability of a 1-atom instantiated by a new constant, or $g>1$ and we consider the conditional probability of a g-atom instantiated by constants already appearing in the evidence.

These conditions held in the binary case (that is, when $r=2$) considered in the previous section. For a general r, we propose the following generalisation of the Binary Sufficientness Postulate:

[10] Note that by the convention from page 31, any partial state description $\Delta(a_1,\ldots,a_m)$ is 1-complete.

The Polyadic Sufficientness Postulate *Let $\Delta(a_1, \dots, a_m)$ be a partial state description as in (50).*

(i) $w(\gamma^1_h(a_{m+1}) \,|\, \Delta)$ depends only m and on the number of times that γ^1_h appears amongst the $\gamma^1_{h_{i_1}}$, $i_1 = 1, \dots, m$.

(ii) Let $1 < g \leq r$ and assume that Δ is $(g-1)$-complete. Let $b_1, \dots, b_g$ be from amongst the $a_1, \dots, a_m$, and such that $\{b_1, \dots, b_g\} \notin C^g$. Assume that $\gamma^g_h(b_1, \dots, b_g) \wedge \Delta$ is consistent. Then

$$w(\gamma^g_h(b_1, \dots, b_g) \,|\, \Delta)$$

depends only on g and on

1. *the number of times that γ^g_h or an equivalent atom appear amongst the $\gamma^g_{h_{i_1,\dots,i_g}}$, $\{a_{i_1}, \dots, a_{i_g}\} \in C^g$.*

2. *the number of times that γ^g_h or an equivalent atom* could have appeared *amongst the $\gamma^g_{h_{i_1,\dots,i_g}}$ in Δ. That is, the number of g-sets $\{a_{i_1}, \dots, a_{i_g}\} \in C^g$, such that*
$$\gamma^g_{h_{i_1,\dots,i_g}}(a_{i_1}, \dots, a_{i_g}) \restriction (g-1)$$
is the $(g-1)$-ary trace of $\gamma^g_f(a_{i_1}, \dots, a_{i_g})$ for some $\gamma^g_f \sim \gamma^g_h$.

3. *s^g_j where $j = j(h)$ is such that $\gamma^g_h \in \Gamma^g_j$; that is, the number of atoms γ^g_f such that $\gamma^g_f \sim \gamma^g_h$ and $\Delta \wedge \gamma^g_f(b_1, \dots, b_g)$ is also consistent.*

It might be hoped and expected that the Polyadic Sufficientness Postulate determines an interesting class of probability functions just like Johnson's Sufficientness Postulate and the Binary Sufficientness Postulate do in the unary and binary cases respectively. A natural class of probability functions satisfying this principle, are the probability functions $w^{\lambda_1,\dots,\lambda_r}$ for $\lambda_g \in (0, \infty]$ determined as follows:

If we refer to the number from (i) above[11] as m_h, then

$$w(\gamma^1_h(a_{m+1}) \,|\, \Delta) \;=\; \frac{m_h + \frac{\lambda_1}{2^q}}{m + \lambda_1}.$$

[11]That is, if m_h is the number of times that γ^1_h appears amongst the $\gamma^1_{h_{i_1}}$.

If $j = j(h)$ and we refer to the numbers[12] from (ii-1) and (ii-2) as n_j^g and v_j^g respectively, then

$$w(\gamma_h^g(b_1, \ldots, b_g) \,|\, \Delta) \;=\; \frac{\frac{n_j^g}{s_j^g} + \frac{\lambda_g}{s^g}}{v_j^g + \lambda_g}$$

(recall that q is the total number of relation symbols in our language and s^g was defined on page 32). We know that in the binary case discussed in the previous section, these are the unique regular probability functions satisfying Ex and the above principle. The general case remains to be further investigated.

4. Conclusion

We based our investigation on the notion of (polyadic) atoms as our central building blocks, since they provide the smallest complete units from which state descriptions (and hence all sentences) can be built. Using this, we were able to propose, first in the binary context and then for general polyadic languages, generalisations of the unary concept of a signature, and principles based on invariance under signatures (BEx, Sgx), independence, instantial relevance and an irrelevance principle generalising Johnson's Sufficientness Postulate, as well as probability functions satisfying these. We have also introduced the more general g-atoms for $g \leq r$ (where r is the arity of the language), and used these to define partial state descriptions. In addition, we were able to completely characterise the probability functions satisfying BEx, BEx+IP and Ex+BIP for binary languages. We have seen that Ex does not imply Sgx and that the signature of a state description is not a determining characteristic of it (up to a permutation of constants) as in the unary case.

This opens the door to many new questions arising from these ideas. For example, 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

[12] That is, n_j^g is the number of times that γ_h^g or an equivalent atom appear amongst the $\gamma^g_{h_{i_1,\ldots,i_g}}$, and v_j^g is the number of times that γ_h^g or an equivalent atom could have appeared amongst the $\gamma^g_{h_{i_1,\ldots,i_g}}$ in Δ.

may define the g-*signature* of a state description for m individuals (where $g \leq m$) analogously to (r-)signatures. We end this paper with some observations regarding these g-signatures, a direction to be further researched.

It is easy to 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 14. 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

$$\begin{array}{ccc} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 2 & 2 & 1 \end{array} \qquad \begin{array}{ccc} 1 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \end{array} \qquad \begin{array}{ccc} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 1 & 1 & 1 \end{array} \qquad \begin{array}{ccc} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{array} \qquad \begin{array}{ccc} 1 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{array} \qquad \begin{array}{ccc} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 2 & 1 & 1 \end{array}$$

are equivalent. Furthermore, the following two are also equivalent:

$$\begin{array}{ccc} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \end{array} \qquad\qquad \begin{array}{ccc} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{array}\,.$$

However,

$$\begin{array}{cccccc} 1 & 1 & 2 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 & 2 & 1 \end{array} \qquad\qquad \begin{array}{cccccc} 1 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 1 \end{array}$$

feature only the above 3-atoms and it can be checked that they have the same 3-signature but not the same 4-signature.

References

[1] Carnap, R., 1963. Replies and systematic expositions. In: Schilpp, P. A. (Ed.), The Philosophy of Rudolf Carnap. Open Court Publishing Company, La Salle, Illinois, pp. 859–1016.

[2] Carnap, R., 1971. A basic system of inductive logic. In: Carnap, R., R.C.Jeffrey (Eds.), Studies in Inductive Logic and Probability I. Univ. California Press, Berkeley, Calif., pp. 33–165.

[3] Carnap, R., 1975. Notes on probability and induction. In: Hintikka, J. (Ed.), Rudolf Carnap, logical empiricist. D. Reidel Publishing Co., Dordrecht-Boston, Mass., pp. 293–324, materials and perspectives, Synthese Library, 73.

[4] Cutland, N., 1995. Loeb measure theory. In: Developments in nonstandard mathematics (Aveiro, 1994). Vol. 336 of Pitman Res. Notes Math. Ser. Longman, Harlow, pp. 151–177.

[5] de Finetti, B., 1974. Theory of probability. Vol. 1. Wiley Classics Library. John Wiley & Sons, Ltd., Chichester.

[6] Gaifman, H., 1964. Concerning measures in first order calculi. Israel J. Math. 2, 1–18.

[7] Johnson, W., 1932. Probability: The deductive and inductive problems. Mind 41, 409–423.

[8] Loeb, P., 1975. Conversion from nonstandard to standard measure spaces and applications in probability theory. Trans. Amer. Math. Soc. 211, 113–122.

[9] Paris, J., Waterhouse, P., 2009. Atom exchangeability and instantial relevance. Journal of Philosophical Logic 38, 313–332.

[10] Paris, J. B., Vencovská, A., 2015. Pure Inductive Logic. The Association of Symbolic Logic series Perspectives in Mathematical Logic. Cambridge University Press.

[11] Ronel, T., 2015. Symmetry principles in polyadic inductive logic. Ph.D. thesis, University of Manchester, Manchester, UK.

[12] Ronel, T., Vencovská, A., 2014. Invariance principles in polyadic inductive logic. Logique et Analyse 228, 541–561.

[13] Royden, H. L., 1968. Real analysis, 2nd Edition. Macmillan Publishing Company, New York.

[14] Vencovská, A., 2006. Binary induction and Carnap's continuum. In: Proceedings of the 7th Workshop on Uncertainty Processing (WUPES). Available at http://www.utia.cas.cz/files/mtr/articles/data/vencovska.pdf, Mikulov, Czech Republic.

[15] Vencovská, A., 2015. Extending Carnap's continuum to binary relations. In: Logic and its applications. Vol. 8923 of Lecture Notes in Computer Science. Springer, pp. 207–217.