THE PRINCIPLE OF SPECTRUM EXCHANGEABILITY WITHIN 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

2009

By
Juergen Landes
Mathematics

Contents

Declaration **6**

Copyright **7**

Acknowledgements **8**

1 General Introduction **9**

2 Mathematical Introduction **12**
- 2.1 Basic Definitions and Notations . . . 12
- 2.2 Introducing the Basic Framework . . . 13
- 2.3 On Equivalence Relations . . . 20
- 2.4 The Spectrum . . . 22

3 Rational Principles **26**
- 3.1 Principles of Exchangeability . . . 26
- 3.2 The Principle of Spectrum Exchangeability is powerful . . . 29

4 The Canonical Representation **32**

5 A first set of Representation Theorems **41**
- 5.1 Introducing $N(\vec{x}, \vec{y}, L)$ . . . 41
- 5.2 Homogeneity and Heterogeneity . . . 47
- 5.3 Forth . . . 47
- 5.4 and back again . . . 50
- 5.5 The $\eta-$Representation . . . 51

6 Sublanguages and (SX) **59**

7 De Finetti-Style Representations **60**
7.1 The Heterogeneous Case 60
7.2 The Homogeneous Case 70

8 Language Invariance **76**

9 The $\eta-$Representation of a $u^{\vec{p}}$ **82**

10 Including Equality **89**
10.1 Extending our Framework 89
10.2 The Principle of Spectrum Exchangeability and Languages containing Equality 93
10.3 Heterogeneity and **(SX=)** 97
10.4 Language Invariance and Equality 98

11 The Paris Conjecture **101**
11.1 The Conjecture 101
11.2 A stronger result? No! 103
11.3 Language Invariance and the Conjecture - Revisited 104
11.4 The Conjecture and Heterogeneous Probabilities Functions 107
11.5 The Paris Conjecture and **(SX=)** 112

12 The Principle of Instantial Relevance **114**

13 Johnson's Sufficientness Principle **118**
13.1 The Principle 118
13.2 **(JSP)** and **(CX)** on a Polyadic Language 120
13.3 **(JSP)** and Equality 122

14 Principles of Conformity **127**
14.1 Preparations and Examples 127
14.2 A first Conformity Principle 130
14.3 **(SX)** implies **(PGC)** 133
14.4 The General Rationale 136
14.5 **(SX)** implies **(GPC)** 137

15 $\#-$Language Invariance **140**

16 Conclusions **145**

Bibliography **147**

A Standard Theorems **150**

Glossary **152**

Abstract

We investigate the consequences of the principle of Spectrum Exchangeability in inductive logic over the polyadic fragment of first order logic.

This principle roughly states that the probability of a possible world should only depend on how the inhabitants of this world behave with respect to indistinguishability. This principle is a natural generalization of exchangeability principles that have long been investigated over the monadic predicate fragment of first order logic. It is grounded in our deep conviction that in the state of total ignorance all possible worlds that can be obtained from each other by basic symmetric transformations should have the same a priori probability.

After first fixing our framework and showing some basic lemmata we prove that the principle of spectrum exchangeability implies several simple principles of exchangeability that are all based on our conviction that the probability functions should be invariant under basic renaming procedures.

We then go on and show several representation theorems for the probability functions satisfying spectrum exchangeability. One of these representation results shows that we can represent the probability of sentences of a polyadic language in terms of the probability of sentences of a unary language. The other main representation result is a de Finetti-style result that shows that we can write every probability function satisfying spectrum exchangeability as an integral over some basic probability function weighted by a de Finetti prior μ.

After that we use the de Finetti representation results to show a representation result for probability functions satisfying language invariance and spectrum exchangeability. Rather surprisingly it turns out that the notion of language invariance allows us to seamlessly extend our framework to the fragment of first order logic containing the equality symbol and the predicate fragment.

Thereafter we study principles that make *inductive assertions*. We investigate the Paris Conjecture and we prove that in some instances the principle of instantial relevance holds for $t-$heterogeneous probability functions. However this principle fails for the completely independent function.

Furthermore we show that the assumption of the principle of constant exchangeability and Johnsons' sufficientness principle leads to only two trivial probability functions satisfying both these two principles.

Declaration

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

Copyright

Acknowledgements

I'm heavily indebted to my supervisor Jeff Paris and his research associate Alena Vencovská for all their time, patience and effort.
This work was funded by the Marie Curie FP6 Early Stage Training Network MATHLOGAPS. Many thanks to all the organizers, especially to Dugald Macpherson.

Chapter 1

General Introduction

We begin our introduction not with the old Greeks but rather with an example. That is due to the fact that the ancient Greeks didn't consider the problems we are interested in here.

Let us assume we find ourselves in a foreign city where we can't make sense of the traffic signs and we observe that the last six cars were all heading west. What should we believe about the seventh car? Will it also be going west, are we in fact in a one-way street?
The mathematical formulation of this problem goes back to Carnap (see [1], [2], [4] and [5]) and Johnson (see [11]) and others who started to consider such problems in the first half of the last century (for a recent collection of essays on the development of the field see [31]). Their mathematical formulation was of the form: Given that we know that a statement γ holds what should be the probability that a statement δ holds? Which can be denoted as $Prob(\delta|\gamma)$.[1]
This is still very imprecise, for instance what do we mean by *should*? One can appeal to rationality and formulate that: "A person or a computer agent X, arguing on the basis of some criteria we have accepted as rational, gives probability p to $Prob(\delta|\gamma)$."

Back with our example where we observed that the first six cars were all heading west. Can we fully rationally argue that we are looking at a one-way street or that we will eventually see a car going the other way? The answers appear to be blatantly *no*. But which probability p should one choose then?
It seems that the answer, if given by a human being, will always depend on the person herself, her background and in this example if she grew up in a neighborhood with

[1] Read $\delta|\gamma$ as "delta given gamma".

lots of one-way streets. Mathematically this is of course highly unsatisfactory as we didn't get a general abstract result that can be verified by anyone understanding the mathematical rules of proof.
Formally we are looking for a probability function, capturing our notion of probability, taking one statement as argument and assigning it a probability.[2] We will follow the traditional path and let our probabilities be in $[0, 1]$, where probability 1 means that an event is going to happen and probability zero stands for an impossible event.
Up to now the statements have mostly been formulated within the predicate fragment of first order logic containing only unary relation symbols. Only recently has the case of the binary fragment been considered (see for instance [21], [22] and [25], an early exception is [14]). We here will consider the full predicate fragment. At a later stage we will extend our framework to also include first order formulae that contain the equality symbol.

As hinted above it has transpired that it is impossible to come up with one single probability function that satisfies everyone's needs. The way forward has been, and still is, to postulate that our probability function satisfies certain properties and then from these assumptions draw conclusions about the chances of an event occurring; and if possible classify all probability functions satisfying the given set of properties.
This approach is most appealing if either the proposed properties, so called *principles of rationality* or *rationales*, are as uncontroversial as possible and widely accepted or the classification yields some desirable results. For instance there might be only few probability functions obtained by that classification or those functions are easily parameterized. In other circumstances it might be helpful that the classification gives a rich class of simple (i.e. easy to handle) probability functions that can be used as building blocks to construct any probability function satisfying the given set of principles.

The main principle we investigate here, that of *Spectrum Exchangeability*, is a principle of symmetry grounded in our conviction that the probability functions should exhibit a high degree of symmetry. For instance we feel that the probability of a sentence φ should be invariant under renaming constants and under renaming relation symbols.
Applied to the example above about the cars in a foreign city this principle entails that

[2] The probability of δ given γ can then be obtained by conditionalization.

the probability is invariant under reversing the direction of travel of all cars. Furthermore the probability shouldn't depend on whether the street goes from east to west or north to south.

Should the reader feel that the principle of spectrum exchangeability postulates too much, he nevertheless has to take into account at least all probability functions that satisfy spectrum exchangeability. We will classify all functions satisfying spectrum exchangeability.

At last we want to caution against the child in the candy store approach. Accepting such principles of rationality too freely quickly leads to disaster as there might be only trivial probability functions that satisfy a given set of such principles (see Chapter 13 for how quickly things can go wrong).

We will spend very little time on philosophical considerations and won't debate at length why one should accept the here presented rationales. We will rather take the *mathematical* route where we assume that *if* some set of principles holds we *then* can prove certain consequences.

Having accepted a principle on general considerations one is then forced to live with those conclusions for better or worse. Should they turn out to be too strongly opposed to what one thinks one has to reject the principle.

Chapter 2

Mathematical Introduction

2.1 Basic Definitions and Notations

Definition 1. Let A, B be sets. We use $A \subset B$ to denote that $a \in A$ implies $a \in B$. That is we allow for $A = B$ when writing $A \subset B$.

Definition 2. Let $\vec{v}$ be a tuple, then denote the number of entries of $\vec{v}$ by $|\vec{v}|$ and let v_i denote the i-th entry of $\vec{v}$. We call $|\vec{v}|$ the *length* of $\vec{v}$.
From time to time we use $\langle$ to denote the beginning of a tuple and we use $\rangle$ to make clear that the tuple ends here. Let $n@t$ be the tuple of length t with all entries equal to n. So from time to time we might write $\langle n@t\rangle$, for example $\langle m@4\rangle = \langle m, m, m, m\rangle$.
We want to state that we allow the entries of our tuples to be real numbers, for example such a tuple is $\langle \frac{1}{7}@5\rangle$.
We consider the *empty tuple*, $\langle\ \rangle$, to be a tuple.

Definition 3. An *array on p of dimension d* is an object that for every $1 \leq i_1, \ldots, i_d \leq p$ contains an entry. We will only consider arrays with entries that are non-negative integers. We call $i_1, \ldots, i_d$ the *running indices*. For our purposes an array is a way to conveniently store and access integers.
For example a $2-$dimensional array is normally referred to as "matrix" and a one dimensional array is often called "tuple" or "vector". If all the entries in an array are either zero or one we say that we have a *zero-one array*.

Definition 4. For $1 \leq d \leq r$ let H_d be a $d-$dimensional array on p where the entries of H_d are in $\{0, 1, \ldots, 2^{m_d} - 1\}$. In the case of $m_d = 0$ we let $H_d := \emptyset$. We'll use the shorthand notation $\mathcal{H}$ for $\langle H_1, \ldots, H_r\rangle$. We'll denote the entry of H_d at the position $j_1, j_2, \ldots, j_d$ by $(H_d)_{j_1,j_2,\ldots,j_d}$.

Let H, G be non-empty $d-$dimensional arrays where H is on p, G on q and $p \leq q$. We say that H is a *subarray of* G, denoted as $H \subset G$ if and only if for all $1 \leq i_1, \ldots, i_d \leq p$ we have $(H)_{i_1,\ldots,i_d} = (G)_{i_1,\ldots,i_d}$. We write $\mathcal{H} \subset \mathcal{G}$ if and only if for all d $H_d \neq \emptyset$ implies that $H_d \subset G_d$.
So in particular every array is a subarray of itself.

Definition 5. Put $\mathbb{N} := \{1, 2, 3 \ldots\}$. Normally the natural numbers begin at 0. For notational convenience we here start with 1.

2.2 Introducing the Basic Framework

We will work with languages L for predicate logic with countably many constant symbols $a_1, a_2, a_3, \ldots,$ finitely many predicate symbols[1] but for the time being without function symbols or equality. The intended interpretation is that the constants *exhaust* the universe. *That is for every individual we want to talk about there is at least one constant representing it.*
From now we will consider the language as given and fixed until clearly stated otherwise. The symbol L, possibly with various annotations, stands from now on for such a language. Let FL denote the well formed formulae of L, SL the sentences of L (i.e. formulae with no free variables) and $QFFL/QFSL$ the quantifier free formulae/sentences of L.

As we are going to see shortly the variables play a side role and we can mainly concentrate on the variable free fragment of L.

Definition 6. Let L be a language as above. The *arity of* L is defined as $r_0(L) := \max\{arity(R) \mid R \in L \text{ and } R \text{ is a relation symbol}\}$.
Let L_d be the d-ary fragment of L. Also let $m_d(L) := |\{R \in L|\ arity(R) = d\}|$ be the number of d-ary relations symbols in L. If there can't be any confusion the explicit mention of L as an argument will be dropped, we'll simply write r_0 and m_d.
A language is *purely unary* if and only if all relation symbols are unary. That is $r_0 = 1$, in other words $L = L_1$. Similarly a language is *purely* $d-$*ary* if all relation symbols in L have arity d. So in this case $L = L_d$ and $r_0 = d$.

Definition 7. From now on we will assume that for all $1 \leq j \leq r_0(L)$ that the relation symbols in L_d are $R_{d,1}, \ldots, R_{d,m_d}$. Whenever we write $R \in L$ or $R \in L_d$ we mean that R is one of these symbols.

[1] of non-zero arity

Definition 8. An *atomic formula of* L is a sentence in $QFSL$ that doesn't contain $\neg, \wedge$ nor $\vee$. That is it is of the form $R\vec{a}$ for some tuple of constants $\vec{a} = \langle a_{i_1}, \ldots, a_{i_{arity(R)}} \rangle$ for some relation symbol $R \in L$. We let AFL denote the set of *atomic formulae of* L. A *literal* is either an atomic formula or the negation of an atomic formula.

Definition 9. From time to time it will be convenient to denote $R\vec{a}$ by $R^1\vec{a}$ and $\neg R\vec{a}$ by $R^0\vec{a}$. We use $\bigwedge_{\vec{a}\in A} \pm R\vec{a}$ to indicate that for every tuple $\vec{a} \in A$ we either have $R^0\vec{a}$ or $R^1\vec{a}$ as a conjunct in $\bigwedge_{\vec{a}\in A} \pm R\vec{a}$. But never $R^0\vec{a}'$ **and** $R^1\vec{a}'$ as conjuncts in $\bigwedge_{\vec{a}\in A} \pm R\vec{a}$ for the same tuple $\vec{a}' \in A$.
Similarly define $\bigwedge_{R\in S} \pm R\vec{a}$ for $|\vec{a}| = d$ and $S \subset L_d$.

Definition 10. Let $L_d \neq \emptyset$. *A* $d-$*atom on* p is a formula in $QFSL$ of the form:

$$\bigwedge_{j=1}^{m_d} \pm R_{d,j} a_{i_1} \ldots a_{i_d} \tag{2.1}$$

where $\{i_1, \ldots, i_d\} \subset \{1, \ldots, p\}$ is fixed.
Every $d-$atom[2] on a fixed tuple of constants can be represented by a number between 0 and $2^{m_d} - 1$. Simply denote the choices whether a literal appears positive or negative by one respectively zero and think of this list of zeros and ones as a number in binary. For example the $d-$atom $\bigwedge_{j=1}^{m_d} \neg R_{d,j}\vec{a}$ is represented by 0 and $\bigwedge_{j=1}^{m_d} R_{d,j}\vec{a}$ by $2^{m_d} - 1$.

Definition 11. *State descriptions on* p are formulae in $QFSL$ of the form:

$$\bigwedge_{d=1}^{r_0} \bigwedge_{j=1}^{m_d} \bigwedge_{i_1,\ldots,i_d=1}^{p} \pm R_{d,j} a_{i_1} \ldots a_{i_d} \quad . \tag{2.2}$$

For example if L contains three relation symbols where $R_{1,1}, R_{1,2}$ and $R_{3,1}$ then the following formula is a state description on 2

$$\begin{aligned}
&R_{1,1}a_1 \wedge R_{1,1}a_2 \wedge \neg R_{1,2}a_1 \wedge R_{1,2}a_2 \wedge \\
&R_{3,1}a_1a_1a_1 \wedge R_{3,1}a_1a_1a_2 \wedge \neg R_{3,1}a_1a_2a_1 \wedge R_{3,1}a_1a_2a_2 \wedge \\
&\neg R_{3,1}a_2a_1a_1 \wedge \neg R_{3,1}a_2a_1a_2 \wedge \neg R_{3,1}a_2a_2a_1 \wedge \neg R_{3,1}a_2a_2a_2 \quad .
\end{aligned}$$

Let $SD_p(L)$ be the set of state descriptions on p and put $SD(L) := \bigcup_{p=1}^{\infty} SD_p(L)$. The argument L will be dropped when there can't be any confusion.

[2]That we omit to specify on which $p \in \mathbb{N}$ the $d-$atom is means that it is irrelevant for current considerations.

Every state description on p over L_d can be represented by an array H_d on p of dimension d with entries in $\{0, \dots, 2^{m_d} - 1\}$. By our convention in definition 4 on page 12 for empty L_d we have $H_d = \emptyset$.
Every state description α on p over the whole language L can be represented by a tuple of arrays $\mathcal{H} = \langle H_1, H_2, \dots, H_{r_0} \rangle$ where the first of element of this tuple H_1 represents the state description on p over L_1, the second element H_2 represents the binary part and so on. In general H_i is an $i-$dimensional array on p with entries in $\{0, \dots, 2^{m_i} - 1\}$. We call $\mathcal{H}$ the *array representation of* α. In the example above $m_1 = 2, m_2 = 0, m_3 = 1$, $H_1 = \langle 2, 3 \rangle$, $H_2 = \emptyset$ and H_3 is a $3-$dimensional zero-one array on 2 with $(H_3)_{1,1,1} = (H_3)_{1,1,2} = (H_3)_{1,2,2} = 1$. All other entries in H_3 are zero. From now on we will use the notion of state description and its array representation interchangeably. Before carrying on we'll give yet another way of looking at state descriptions.
For every $1 \le d \le r_0$ and every relation symbol $R \in L_d$ we look at

$$\psi := \bigwedge_{i_1, \dots, i_d = 1}^{p} \pm R a_{i_1} \dots a_{i_d}$$

and note that we can represent this formula by one $d-$dimensional zero-one array on p. The entry in the array at position $i_1, \dots, i_d$ is zero if $\psi \vDash \neg R a_{i_1} \dots a_{i_d}$ and one if $\psi \vDash R a_{i_1} \dots a_{i_d}$.
Then overall we can represent a state description by m_1 many $1-$dimensional, m_2 many $2-$dimensional up to $m_{r_0}-$many r_0-dimensional zero-one arrays on p. In our example the $1-$dimensional arrays are $\langle 1, 1 \rangle$ representing $R_{1,1} a_1 \wedge R_{1,1} a_2$ and $\langle 0, 1 \rangle$ representing $\neg R_{1,2} a_1 \wedge R_{1,2} a_2$.
Clearly all three ways of looking at state descriptions are easily interchangeable. If we want to stress that we want to work with this third formulation using zero-one arrays we write $\alpha \in SD_p^{01}$ or $\mathcal{H} \in SD_p^{01}$.

We duly note that every state description on p decides every atomic formula in $QFSL$ containing only the first p constants and that every state description is consistent.[3]

These notions of state descriptions treat different constants differently, for instance a_1 is in every state description and a_5 is not. For example a_5 does not feature in any $\alpha \in SD_4$. As eluded to in the introduction we will later on assume that the principle

[3] We say that a formula ψ is consistent if and only if there is a first order structure $\mathcal{A}$ and evaluation e such that $\mathcal{A} \vDash e(\psi)$.

of spectrum exchangeability holds. This principle implies the principle of constant exchangeability (see page 27). The latter states that the constants are interchangeable. Hence the assumption that a state description is always on $a_1, \ldots, a_p$ and not on different $\{a_{i_1}, \ldots, a_{i_p}\}$ does not pose a real restriction to us.

For later use we make the following

Remark 1. Let $L_d \neq \emptyset$. Then the number of $d-$atoms on fixed $a_{i_1}, \ldots, a_{i_d}$ is 2^{m_d}. The number of literals in an $\alpha \in SD_p(L_d)$ equals $m_d \cdot p^d$. Furthermore $(2^{m_d})^{(p^d)} = |SD_p(L_d)|$ and so $|SD_p(L)| = \prod_{d=1}^{r_0}(2^{m_d})^{(p^d)}$.
For a fixed $\beta \in SD_p(L)$ we have for $q > p$ that $|\{\alpha \in SD_q(L)|\alpha \vDash \beta\}| = \frac{|SD_q(L)|}{|SD_p(L)|}$.

Definition 12. A map $w : SL \to [0,1]$ is called a *probability function on* L if and only if the three following conditions hold for all $\theta, \varphi, \exists x\psi(x) \in SL$:
(P1) If $\models \theta$ then $w(\theta) = 1$.
(P2) If $\models \neg(\theta \wedge \varphi)$ then $w(\theta \vee \varphi) = w(\theta) + w(\varphi)$.
(P3) $w(\exists x\psi(x)) = \lim_{m\to\infty} w(\bigvee_{i=1}^{m} \psi(a_i))$.

The third condition (P3) captures the fact that the constants *exhaust* the universe.

If we identify belief with willingness to bet then the *Dutch Book Argument* (see for instance [23] pp. 19) forces us to identify belief with probability. That is any function reflecting our notion of belief has to satisfy the three properties of *probability* above. The argument shows that any other belief assignment entails a certain loss, i.e. Dutch Book.

From now on w, possibly with various annotations, denotes a probability function. There are some immediate conclusions to be drawn that are in line with our intuition as to how a probability function should behave.

Lemma 1. For $\varphi, \psi \in SL$ and every probability function w on L it holds that

1. $w(\neg\varphi) = 1 - w(\varphi)$.
2. If $\vDash \varphi$ then $w(\neg\varphi) = 0$.
3. If $\vDash \varphi \to \psi$ then $w(\varphi) \leq w(\psi)$.
4. If $\vDash \varphi \leftrightarrow \psi$ then $w(\varphi) = w(\psi)$.
5. $w(\varphi \vee \psi) = w(\varphi) + w(\psi) - w(\varphi \wedge \psi)$.

Proof. See for instance [23] pp. 10 or [24]. Both references also contain some background on the Dutch Book Argument mentioned above. □

So logically equivalent formulae have the same probability.
These observations above will frequently be used and at times without further mention.

Gaifman has shown in [8] the following very helpful fact

Theorem 1. *Any function $v : QFSL \to [0,1]$ satisfying (P1) and (P2) on $QFSL$ extends uniquely to a probability function on L.*

From now on we will concentrate mainly on $QFSL$. The idea is that we can always use Gaifmans' theorem to obtain the probability of sentences containing quantifiers. And since we are mainly interested in the probability of sentences variables all but disappear in the following.

Definition 13. Let $\psi \in SL$ then define $const(\psi) := \max\{k \mid a_k \in \psi\}$.

Lemma 2. Let $p, q \in \mathbb{N}$ be such that $p \leq q$. Then for all consistent conjunctions of literals $\psi \in QFSL$ with $const(\psi) = p$ we have

$$\models \psi \leftrightarrow \bigvee_{\substack{\beta \in SD_q \\ \beta \models \psi}} \beta \ . \tag{2.3}$$

Proof. Let $\Gamma_q := \{\gamma \in AFL \mid const(\gamma) \leq q,\ \psi \not\models \gamma,\ \psi \not\models \neg\gamma\}$. Then

$$\models \psi \leftrightarrow \ \Big(\psi \ \wedge \bigwedge_{\gamma \in \Gamma_q} (\gamma \vee \neg\gamma)\Big)$$
$$\models \psi \leftrightarrow \bigvee_{\substack{\beta \in SD_q \\ \beta \models \psi}} \beta$$

by the fact that the state descriptions have to decide all atomic formulae and the distribution laws. □

Lemma 3. Let $p, q \in \mathbb{N}$ be such that $p \leq q$. Then for all consistent $\psi \in QFSL$ with $const(\psi) = p$ we have

$$\models \psi \leftrightarrow \bigvee_{\substack{\beta \in SD_q \\ \beta \models \psi}} \beta \ .$$

Proof. We use the disjunctive normal form[4] theorem [see Appendix, Theorem 25 on page 150] to obtain the DNF of ψ as $\vDash \psi \leftrightarrow \bigvee_i \alpha_i$ where all the α_i are consistent conjunction of literals in $QFSL$ with $const(\alpha_i) \leq p$. Furthermore all the α_i imply ψ. Hence by applying the lemma above to the α_i we find $\vDash \alpha_i \leftrightarrow \bigvee_{\substack{\beta \in SD_q \\ \beta \vDash \alpha_i}} \beta$ and hence $\vDash \psi \leftrightarrow \bigvee_{\substack{\beta \in SD_q \\ \beta \vDash \psi}} \beta$. □

Lemma 4. Let $f : SD \to \{r \in \mathbb{R} \mid r \geq 0\}$ be such that $\sum_{\alpha \in SD_1} f(\alpha) = 1$ and such that for all p and all $\beta \in SD_p$ we have

$$f(\beta) = \sum_{\substack{\gamma \in SD_{p+1} \\ \gamma \vDash \beta}} f(\gamma) \ .$$

Then f extends uniquely to a probability function w on $QFSL$.

Proof. We will first assume that such an extension w exists from which we'll derive that for every $\psi \in QFSL$ there's only one possible value we can give to $w(\psi)$. It then remains to show that this indeed does define a probability function.
First of all since logically equivalent formulae have the same probability we have to have $w(\chi) = 0$ for all contradictory formulae χ and $w(\rho) = 1$ for all tautologies ρ.
Now suppose that $\psi \in QFSL$ is neither contradictory nor a tautology and assume that $const(\psi) = p$. Then as seen above $\vDash \psi \leftrightarrow \bigvee_{\substack{\beta \in SD_p \\ \beta \vDash \psi}} \beta$. So w has to satisfy that $w(\psi) = w(\bigvee_{\substack{\beta \in SD_p \\ \beta \vDash \psi}} \beta)$. We put $A_p := \{\beta \in SD_p \mid \beta \vDash \psi\}$ and so $w(\psi) = w(\bigvee_{\beta \in A_p} \beta)$.
But since for different $\beta, \beta' \in A_p$ there is an atomic formula in $QFSL$ that is decided in different ways by β and β' we have $\models \neg(\beta \wedge \beta')$ and hence $w(\beta \wedge \beta') = 0$.

[4] From now on just DNF

Next suppose we enumerated the $\beta \in A_p$ as $\beta_1, \ldots, \beta_n$ for some $n \in \mathbb{N}$. Then extension w of f has to satisfy

$$w(\psi) = w(\bigvee_{\beta \in A_p} \beta) = w(\bigvee_{i=1}^{n} \beta_i) = w(\beta_1 \vee \bigvee_{i=2}^{n} \beta_i) \tag{2.4}$$

$$= w(\beta_1) + w(\bigvee_{i=2}^{n} \beta_i) + w(\beta_1 \wedge \bigvee_{i=2}^{n} \beta_i) \tag{2.5}$$

$$= w(\beta_1) + w(\bigvee_{i=2}^{n} \beta_i) + w(\bigvee_{i=2}^{n} \beta_1 \wedge \beta_i) \tag{2.6}$$

$$= w(\beta_1) + w(\bigvee_{i=2}^{n} \beta_i) \tag{2.7}$$

$$= \sum_{\beta \in A_p} w(\beta) = \sum_{\substack{\beta \in SD_p \\ \beta \models \psi}} w(\beta) = \sum_{\substack{\beta \in SD_p \\ \beta \models \psi}} f(\beta) \ . \tag{2.8}$$

(2.4) and (2.6) hold because of (4) of Lemma 1. (2.5) follows from (5) of Lemma 1. (2.7) is true since for $i \geq 2$ $\beta_1 \wedge \beta_i$ is inconsistent. Hence $\bigvee_{i=2}^{n}(\beta_1 \wedge \beta_i)$ is inconsistent. (2.8) is by induction on n and the definition of A_p.

So we have assigned every $\psi \in QFSL$ a unique value under w. Hence we have shown that there is at most one probability function extending f. In the next part we will prove that the function w defined by (2.8) is indeed a probability function.

Clearly w maps $QFSL$ to the non-negative reals and w satisfies (P1). Furthermore we have since every $\gamma \in SD_{p+1}$ implies exactly one $\delta \in SD_p$ that

$$\sum_{\delta \in SD_p} f(\delta) = \sum_{\delta \in SD_p} \sum_{\substack{\gamma \in SD_{p+1} \\ \gamma \models \delta}} f(\gamma) = \sum_{\gamma \in SD_{p+1}} f(\gamma) \ .$$

Since $\sum_{\alpha \in SD_1} f(\alpha) = 1$ by our assumption we have $1 = \sum_{\delta \in SD_p} f(\delta)$ for all p. Furthermore for all p, i and $\alpha \in SD_p$ we find

$$f(\alpha) = \sum_{\substack{\beta \in SD_{p+i} \\ \beta \models \alpha}} f(\beta) \ . \tag{2.9}$$

So the value we assigned $w(\psi)$ in (2.8) does not change if we replace p by $p + i$.

For any $\psi \in QFSL$ with $const(\psi) = p$ that is not contradictory nor a tautology we have

$$0 \leq w(\psi) = \sum_{\substack{\beta \in SD_p \\ \beta \vDash \psi}} f(\beta) \leq \sum_{\beta \in SD_p} f(\beta) = \sum_{\beta \in SD_1} f(\beta) = 1 \ .$$

It only remains to check that (P2) holds. Consider therefore $\psi, \chi \in QFSL$ such that $\vDash \neg(\psi \wedge \chi)$. If ψ or χ is a contradiction or a tautology then there is nothing to prove. Put $p := \max\{const(\psi), const(\chi)\}$ then

$$\vDash (\psi \vee \chi) \leftrightarrow \Big(\bigvee_{\substack{\alpha \in SD_p \\ \alpha \vDash \psi}} \alpha \vee \bigvee_{\substack{\beta \in SD_p \\ \beta \vDash \chi}} \beta \Big)$$

$$\vDash (\psi \vee \chi) \leftrightarrow \bigvee_{\substack{\alpha \in SD_p \\ \alpha \vDash (\psi \vee \chi)}} \alpha \ .$$

Then $w(\psi \vee \chi) = w(\bigvee_{\substack{\alpha \in SD_p \\ \alpha \vDash (\psi \vee \chi)}} \alpha)$. Since an $\alpha \in SD_p$ can imply at most one of ψ or χ and since the conjunction of two different state descriptions in SD_p is inconsistent we find using (2.9) to obtain (2.11)

$$w(\psi \vee \chi) = w(\bigvee_{\substack{\alpha \in SD_p \\ \alpha \vDash (\psi \vee \chi)}} \alpha) = \sum_{\substack{\alpha \in SD_p \\ \alpha \vDash (\psi \vee \chi)}} f(\alpha) = \sum_{\substack{\alpha \in SD_p \\ \alpha \vDash \psi}} f(\alpha) + \sum_{\substack{\beta \in SD_p \\ \beta \vDash \chi}} f(\beta) \tag{2.10}$$

$$= w(\psi) + w(\chi) \ . \tag{2.11}$$

□

We want to note that setting $w(\psi) = \sum_{\substack{\beta \in SD_p \\ \beta \vDash \psi}} f(\beta)$ as in (2.8) also gives the right result if ψ is a tautology or a contradiction. In case of a tautology all $\beta \in SD_p$ imply ψ and in case of $\vDash \neg\psi$ no state description implies ψ.

The last lemma shows that every probability function on L is completely determined by its values on $SD(L)$. We can henceforth identify a probability function on $SD(L)$ with the unique probability function it induces on SL.

2.3 On Equivalence Relations

Later on we will frequently deal with equivalence relations. We devote this section to set up the the basic notions and definitions.

Definition 14. Let ER_p be the *set of equivalence relations on* $\{1, \ldots, p\}$. We say that $E \in ER_p$ is *an equivalence relation on* p.
Suppose $E \in ER_p$ and $1 \leq i, k \leq p$. Denote by $E(i \mapsto k, k \mapsto i)$ the equivalence relation obtained from E by having i and k swap classes. If i and k are equivalent over E then $E = E(i \mapsto k, k \mapsto i)$.
We denote by $|E|$ the number of classes of E which will be called its *size*. We let $ER_p^t := \{E \in ER_p \mid |E| = t\}$ that is the equivalence relations on p with t classes.
We write Eij to denote that i and j are equivalent over E, i.e. there are in the same class of E. To ease the reading we'll occasionally put parenthesis around i or j. So for example we'll write $E(i)j$.

Definition 15. Let $E \in ER_p$ and let $\{i_1, \ldots, i_k\} \subset \{1, \ldots, p\}$. Then define the *restriction* of E to $i_1, \ldots, i_k$ as the equivalence relation $E \upharpoonright_{i_1,\ldots,i_k}$ on $\{i_1, \ldots, i_k\}$ which holds for $l, j \in \{i_1, \ldots, i_k\}$ if and only if Elj. If $\{i_1, \ldots, i_k\} = \{1, \ldots, q\}$ for some $q \leq p$ we simply write $E \upharpoonright_q$.
Let $E \in ER_p^t$ and let $E_1, \ldots E_t$ be the t classes of E. From now on we assume that the classes $E_1, \ldots, E_t$ are ordered by the following condition: for $i < l$ we have that $\min\{e \in E_i\} < \min\{e \in E_l\}$; if not clearly stated otherwise. We write $E = \langle E_1, \ldots, E_t \rangle$.

Definition 16. Let $E, D \in ER_p$. We say the D is a *refinement of* E if and only if every class E_i of E is the union of some classes D_j of D. That D is a refinement of E will be denoted by $E \leq D$. Then necessarily $|E| \leq |D|$.
So $E \leq D$ holds whenever for equivalent i, k over D it follows that i and k are also equivalent over E.

Definition 17. Let $E \in ER_p^t$. Then put $v_s := \min\{r | r \in E_s\}$ for $1 \leq s \leq t$. We define the *very important elements of* E to be the set of these elements $V(E) := \{v_1, v_2, \ldots, v_t\}$.
We use $v(i)$ to denote the very important element that is in the same class as i. We call $v(i)$ the *representative of* i. For a tuple of constants $\vec{a} := a_{i_1}, \ldots, a_{i_d}$ with $1 \leq i_1, \ldots, i_d \leq p$ we write $v(\vec{a})$ for $a_{v(i_1)}, \ldots, a_{v(i_d)}$.
Define a function $vip : \{1, \ldots, p\} \rightarrow \{1, \ldots, t\}$ via $vip(i) := n$ if and only if $i \in E_n$, equivalently $v(i) = v_n$. Again we put for a tuple of constant $\vec{a}$ as above $vip(\vec{a}) := a_{vip(i_1)}, \ldots, a_{vip(i_d)}$. Note for later use that $vip(v(i)) = vip(i)$.
If there can be some confusion we'll write v_E and vip_E to make clear that we work with respect to an equivalence relation E.

In later applications we might prefer to have our equivalence relations on $\{a_1, \ldots, a_p\}$ rather than on $\{1, \ldots, p\}$. We will choose whatever is more convenient at the time.

Definition 18. Let $E \in ER_p^t$ with $E = \langle E_1, \ldots, E_t \rangle$. Now we arrange the classes of E such that the biggest class comes first, then the second biggest class and so on. In case some classes have the same size order by smallest element; that is the class with the smallest element comes first. We denote by $\mathcal{S}(E)$ the result of this reshuffle. Denote by x_i the size of the $i-$th class in this order.
Then the *spectrum of* E is then defined as the tuple $\langle x_1, \ldots, x_t \rangle$. We write $\overrightarrow{\mathcal{S}(E)}$ to denote the spectrum of E. Please note that the reshuffle now implies that $x_i \geq x_{i+1}$ for all $1 \leq i \leq t-1$.

Definition 19. Let E be an equivalence relation. We say that i is a singleton over E if and only if there is a class E_k of E such that $E_k = \{i\}$.

To illustrate our most recent definitions we give the following

Example 1. Let $E \in ER_8^4$ such that $1 \sim_E 5$, $2 \sim_E 4 \sim_E 8$, $3 \sim_E 7$ and 6 is a singleton over E. Then $E_1 = \{1, 5\}$, $E_4 = \{6\}$ and $E = \langle \{1, 5\}, \{2, 4, 8\}, \{3, 7\}, \{6\} \rangle$ and $\mathcal{S}(E) = \langle \{2, 4, 8\}, \{1, 5\}, \{3, 7\}, \{6\} \rangle$. The spectrum $\overrightarrow{\mathcal{S}(E)}$ of E is $\langle 3, 2, 2, 1 \rangle$ and $V(E) = \{1, 2, 3, 6\}$. Furthermore $v_E(8) = 2$, $v_E(6) = 6$ and $vip_E(6) = 4$ since $6 \in E_4$.
$E \restriction_4 = \langle \{1\}, \{2, 4\}, \{3\} \rangle$ and $E \restriction_{\{1,2,5,6,7,8\}} = \langle \{1, 5\}, \{2, 8\}, \{6\}, \{7\} \rangle$.
Let $D \in ER_8^3$ such that $1 \sim_D 5 \sim_D 6$, $2 \sim_D 4 \sim_D 8$, $3 \sim_D 7$. Then $D < E$ and for instance $vip_D(6) = 1$.

2.4 The Spectrum

Definition 20. Let $\varphi \in SL$. Then denote by $\varphi(a_i \mapsto a_j)$ the result of replacing every occurrence of a_i by a_j in φ. By $\varphi(a_i \mapsto a_j, a_j \mapsto a_i)$ we denote the result of swapping a_i and a_j in φ.

Definition 21. Consider a state description $\alpha \in SD_p$ and its array representation $\mathcal{H}$. We'll define an equivalence relation $S(\alpha)$[5] on p. We call $S(\alpha)$ the *spectral equivalence relation of* α.
There are several ways of defining $S(\alpha)$ all of which are equivalent as we will shortly

[5] or $S(\mathcal{H})$: as mentioned we won't distinguish between a state description α and its representation in array formulation $\mathcal{H}$.

see. Here's a list of possible definitions.
We say that over α or $S(\alpha)$ i and j are equivalent if and only if

1. a_i and a_j are indistinguishable in α.
   That is: For all pairs of literals λ, λ' with $const(\lambda), const(\lambda') \leq p$ such that one can be obtained from the other by replacing some a_i by a_j and simultaneously some a_j by a_i it holds that $\alpha \models \lambda \leftrightarrow \lambda'$.

2. For all relation symbols $R \in L$ and for all tuples $\vec{a}, \vec{a}'$ such that $|\vec{a}| + |\vec{a}'| = arity(R) - 1$[6] and such that the elements of these tuples are in $\{a_1, \ldots, a_p\}$ we have
$$\alpha \models R\vec{a}a_i\vec{a}' \leftrightarrow R\vec{a}a_j\vec{a}' \quad .$$

3. For all relation symbols $R \in L$ and for all tuples $\vec{a}, \vec{a}'$ such that $|\vec{a}| + |\vec{a}'| = arity(R) - 1$ and such that the elements of these tuples are in $\{a_1, \ldots, a_p\}$ we have
$$\alpha \vDash R\vec{a}a_i\vec{a}' \quad \text{if and only if} \quad \alpha \vDash R\vec{a}a_j\vec{a}' \quad .$$

4. For all $\emptyset \neq H_d$ in $\mathcal{H}$ and all tuples $\vec{k}, \vec{k}'$ such that $|\vec{k}| + |\vec{k}'| = d - 1$ and such that the elements of these tuples are in $\{1, \ldots, p\}$ we have
$$(H_d)_{\vec{k}i\vec{k}'} = (H_d)_{\vec{k}j\vec{k}'} \quad .$$

5. For every d and $d-$dimensional zero-one array H_d in $\mathcal{H} \in SD_p^{01}$ and all tuples $\vec{k}, \vec{k}'$ such that $|\vec{k}| + |\vec{k}'| = d - 1$ and such that the elements of these tuples are in $\{1, \ldots, p\}$ we have
$$(H_d)_{\vec{k}i\vec{k}'} = (H_d)_{\vec{k}j\vec{k}'} \quad .$$

6. $\alpha(a_i \mapsto a_j)$ is consistent.

As indicated above; from time to time it is convenient to consider $S(\alpha)$ as an equivalence relation on $\{a_1, \ldots, a_p\}$ instead of $\{1, \ldots, p\}$. For our purposes it doesn't matter much which point of view one takes and we'll use depending on the situation the more convenient notion.

Lemma 5. All ways of defining $S(\alpha)$ are equivalent.

---

[6] According to our convention $\vec{a}, \vec{a}'$ are allowed to be the empty tuple. Indeed they will be both the empty tuple for a unary symbol R.

Proof. Clearly the first three definitions are equivalent and so are the fourth and the fifth definition.
From the definition of the array representation of a state description it follows directly that definition two and five are the same.
We will now show that the last condition follows from the second and that the last condition implies the third.
α is a state description and is hence consistent. Our second formulation shows that we can replace successively all a_i by equivalent a_j[7] in α and have a consistent formula all the way. Hence $\alpha(a_i \mapsto a_j)$ is consistent.
Now suppose that $\alpha(a_i \mapsto a_j)$ is consistent. Consider a literal λ containing a_j such that $\alpha \vDash \lambda$. Let λ' be obtained from λ by replacing one occurrence of a_j by a_i. Hence λ contains a negation symbol if and only if λ' does. Since $\alpha(a_i \mapsto a_j)$ is consistent we have $\alpha \vDash \lambda'$. This was our third definition. □

The last condition might seem counter intuitive since it's not plainly symmetric in i and j. Nevertheless, as we have just seen, it is as good as any of the other possible ways of defining $S(\alpha)$.
The following is already implicit in the text above. Due to its significance we separately state

Definition 22. Let $\alpha \in SD_p$. Then we define the *spectrum of* α to be the spectrum of $S(\alpha)$ denoted by $\overrightarrow{\mathcal{S}(\alpha)}$.

Example 2. Consider the following two-dimensional arrays on 4 each representing a state description over L_2, where L_2 contains 3 relation symbols.

$$H = \begin{matrix} 0 & 2 & 0 & 0 \\ 7 & 0 & 7 & 7 \\ 0 & 2 & 0 & 0 \\ 5 & 1 & 5 & 5 \end{matrix} \qquad G = \begin{matrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{matrix} \qquad I = \begin{matrix} 2 & 1 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 \end{matrix}$$

Then $S(H) = \langle\{1,3\},\{2\},\{4\}\rangle$, $S(G) = \langle\{1\},\{2,4\},\{3\}\rangle$ and
$S(I) = \langle\{1\},\{2\},\{3\},\{4\}\rangle$.

Remark 2. Let $\mathcal{H}, \mathcal{G} \in SD_p$ be such that for all $1 \leq d \leq r_0$ with $m_d \neq 0$ we have $S(H_d) = S(G_d)$.[8] Then $S(\mathcal{H}) = S(\mathcal{G})$ since i and j are equivalent over $S(\mathcal{H})$ if and

[7] equivalent over $S(\alpha)$ that is

[8] $S(H_d)$ is defined as the spectral equivalence relation of H_d which is in $SD_p(L_d)$. Since H_d is a state description over some language $S(H_d)$ is well defined.

only if they are equivalent over $S(H_d)$ for all $1 \leq d \leq r_0$ with $m_d \neq 0$.

Example 3. Let $\mathcal{H} \in SD_p(L)$ be such that $S(H_i) = \langle\{1\}, \{2\}, \ldots, \{p\}\rangle$ for one $1 \leq i \leq r_0$. Then $S(\mathcal{H}) = \langle\{1\}, \{2\}, \ldots, \{p\}\rangle$.
Let $\mathcal{F} \in SD_6$ with $r_0(L) = 4$ such that

$$\begin{aligned} S(F_1) &= \langle\{1,2,3,4,5,6\}\rangle, \quad S(F_2) = \langle\{1,2,3\},\{4,5,6\}\rangle, \\ S(F_3) &= \langle\{1,6\}\{2,3,4,5\}\rangle, \quad S(F_4) = \langle\{1,2,3,5\}\{4,6\}\rangle \ . \end{aligned}$$

Then $S(\mathcal{F}) = \langle\{1\}, \{2,3\}, \{4\}, \{5\}, \{6\}\rangle$.

Definition 23. Let X denote the *set of all spectra*. Define furthermore

$$X_p := \{\vec{x} \in X \mid \sum_{i=1}^{|\vec{x}|} x_i = p\} \qquad X_{\leq p} := \{\vec{x} \in X \mid \sum_{i=1}^{|\vec{x}|} x_i \leq p\} \tag{2.12}$$

$$X_p^t := \{\vec{x} \in X_p \mid |\vec{x}| = t\}, \qquad X^{\leq t} := \{\vec{x} \in X \mid |\vec{x}| \leq t\} \tag{2.13}$$

and any combination of them, for instance $X_p^{\leq t} := \{\vec{x} \in X_p \mid |\vec{x}| \leq t\}$. We say that an $\vec{x} \in X_p^t$ is a *spectrum on* p *of length* t.
Let SD_p^t denote the set of all state descriptions in SD_p with spectrum of length t. Similarly define $SD_p^{\leq t}$ and $SD_p^{\geq t}$. If $\alpha \in SD_p^t$ then $S(\alpha) \in ER_p^t$ and hence the spectrum of α is in X_p^t.

To acquaint the reader with our latest definitions we give the following

Example 4.

$$X_1 = \{\langle 1\rangle\}, \ \ X_2 = \{\langle 2\rangle, \langle 1,1\rangle\}, \ \ X_t^t = \{\langle 1@t\rangle\} \tag{2.14}$$

$$X_4 = \{\langle 4\rangle, \langle 3,1\rangle, \langle 2@2\rangle, \langle 2,1,1\rangle, \langle 1@4\rangle\}, \tag{2.15}$$

$$X_t^2 = \bigcup_{1 \leq i \leq \frac{t}{2}} \{\langle t-i, i\rangle\}, \ \ X_t^{t-1} = \{\langle 2, 1@t-2\rangle\}, \tag{2.16}$$

$$X_t^1 = \{\langle t\rangle\} \ . \tag{2.17}$$

Chapter 3

Rational Principles

3.1 Principles of Exchangeability

All principles in this section are proposed for the case of total ignorance. That is we do not hold any preconceived ideas of the state of the world. A computer scientist would say that our knowledge base is empty.

That is definitely the easiest case but also the one where we feel that there is some chance to get a good proportion of people to agree that the here proposed principles are not completely arbitrary and that it does make sense to consider them. A computer scientist might add that all algorithms start in the state of complete ignorance.

Principle of **Spectrum Exchangeability** **(SX)**

Let $\alpha, \beta \in SD$ have the same spectrum. Then

$$w(\alpha) = w(\beta) \ . \tag{3.1}$$

Note that for $p \neq q$ with $\alpha \in SD_p$ and $\beta \in SD_q$ we always have that $\overrightarrow{\mathcal{S}(\alpha)} \neq \overrightarrow{\mathcal{S}(\beta)}$ since the spectrum of α is on p and that of β is on q.

If we accept **(SX)** we can unambiguously write $w(\vec{x})$ for $\vec{x} \in X$ meaning the probability of one/any state description with spectrum $\vec{x}$. The principle of spectrum exchangeability entails that all state description with spectrum $\vec{x}$ have the same probability and so $w(\vec{x})$ is well defined.

Principle of **Regularity** **(REG)**

If $\psi \in QFSL$ is satisfiable then $w(\psi) > 0$.

Remark 3. Let w be a probability function on L and suppose that $\psi \in QFSL$. Then it follows from **(REG)** that: $w(\psi) = 0 \iff \models \neg\psi$.
So the probability of a state description would always be greater than zero if we accepted **(REG)**.
(REG) means that unless some statement is contradictory we can't rule out completely that it holds. So **(REG)** says that "Nothing is Impossible; as long as it is not outright contradictory."

Principle of **Strong Negation (SN)**
For all $\varphi, \psi \in QFSL$, if P is any relation symbol of L and ψ is obtained from φ by replacing each occurrence of P in φ by $\neg R$ then $w(\varphi) = w(\psi)$.

Example 5. Let $\varphi = Pa_1 \wedge \neg Pa_2 \wedge Ra_1$ and $\psi = \neg Pa_1 \wedge \neg\neg Pa_2 \wedge Ra_1$. Then **(SN)** implies that φ and ψ have the same probability.

Principle of **Weak Negation (WN)**
For all $\varphi, \psi \in QFSL$, if ψ is obtained from φ by replacing every occurrence of every relation symbol in φ by its negation then $w(\varphi) = w(\psi)$.

Clearly strong negation is in general stronger than weak negation. If L contains only one relation symbol then **(SN)** and **(WN)** are equally strong.

Remark 4. Let w satisfy **(WN)**. Then for every literal λ we have $w(\lambda) = \frac{1}{2}$.
That is so since $1 - w(\lambda) = w(\neg\lambda) \stackrel{\textbf{(WN)}}{=} w(\lambda)$.

Principle of **Constant Exchangeability (CX)**
For $\varphi, \psi \in QFSL$, if ψ is obtained from φ by replacing a_i in φ by a_q where a_q is not in φ then $w(\varphi) = w(\psi)$.

This principle states that the probability should be invariant under renaming of individuals. It hence expresses "What's in a name?"
Repeated application of the principle of constant exchangeability shows that for constants $a_k, a_j \in \varphi$ that $w(\varphi) = w(\varphi(a_k \mapsto a_j, a_j \mapsto a_k))$. Subsequently we see that **(CX)** implies that w is then not only invariant under transpositions of constants but also under permutations of constants.

Principle of **Predicate Exchangeability (PX)**
Let $\varphi, \psi \in QFSL$ and let P, Q be predicate symbols of the same arity. If ψ is obtained

from φ by replacing every occurrence of P by Q and simultaneously all occurrences of Q by P, then $w(\varphi) = w(\psi)$.

Definition 24. The group of permutations on d elements will be denoted by S^d.[1]

We will state the next principle only for state descriptions. This poses no real restriction and allows us to state the principle without introducing any new notation.

Principle of **Atom Exchangeability (AX)**
Let $\mathcal{H}, \mathcal{G} \in SD_p$ such that there is a $1 \le d \le r_0$ and a $\sigma \in S^{2^{m_d}}$ such that $\mathcal{G}$ can be obtained from $\mathcal{H}$ by replacing $(H_d)_{i_1,\dots,i_d}$ by $\sigma((H_d)_{i_1,\dots,i_d})$ for all $1 \le i_1, \dots, i_d \le p$. Then $w(\mathcal{H}) = w(\mathcal{G})$.

It is easy to see that **(AX)** implies **(PX)** and **(SN)**.
On a purely unary language assuming **(SX)** is equivalent to assuming **(CX)** and **(AX)**. **(SX)** on an arbitrary polyadic language is hence a straight forward generalization of previously well studied rationales. We consider this and the fact that **(SX)** allows de Finetti representations (see Chapter 7) major motivations to investigate **(SX)**. A further motivation will emerge in Chapter 10.

Definition 25. Let $\vec{a} = \langle a_{i_1}, \dots, a_{i_d}\rangle$ be any tuple of constants of length d and let $\sigma \in S^d$. Then put $\sigma(\vec{a}) := \langle a_{i_{\sigma(1)}}, \dots, a_{i_{\sigma(d)}}\rangle$. So $\sigma(\vec{a})$ affects the order of the entries of $\vec{a}$. For example if $\vec{a} = \langle a_3, a_5, a_2, a_1, a_7, a_7\rangle$ and $\sigma \in S^6$ is the transposition of 2 and 4 then $\sigma(\vec{a}) = \langle a_3, a_1, a_2, a_5, a_7, a_7\rangle$.

Principle of **Variable Exchangeability (VX)**
Let $P \in L_d$ and let $\sigma \in S^d$. Suppose $\chi, \psi \in QFSL$ are such that χ can be obtained from ψ by replacing $P\vec{a}$ by $P\sigma(\vec{a})$ for all tuples $\vec{a}$. Then $w(\chi) = w(\psi)$.

For a binary relation symbol P in L_2 **(VX)** implies for $\varphi \in QFSL$ not containing P that $w(\varphi \wedge \bigwedge_{i,k=1}^{p} P^{\epsilon_{ik}} a_i a_k) = w(\varphi \wedge \bigwedge_{i,k=1}^{p} P^{\epsilon_{ik}} a_k a_i)$.
In the array formalism **(VX)** implies that the probability of a state description $\mathcal{H} \in SD_p^{01}$ is invariant under reordering the running indices of one of its arrays.

[1] Try not to be confused by $S(\alpha)$ and S^d. $S(\alpha)$ and later on $S(c)$ are equivalence relations whereas S^d is a group.

3.2 The Principle of Spectrum Exchangeability is powerful

Lemma 6. **(SX)** implies **(SN)**, **(PX)**, **(VX)**, **(AX)** and **(CX)**.

Proof. Let $\psi \in QFSL$ and $const(\psi) = p$. Then $\models \psi \leftrightarrow \bigvee_{\substack{\alpha \in SD_p \\ \alpha \models \psi}} \alpha$ as seen in Lemma 3 on page 17.
Now let ψ' be the result of applying **(SN)**, **(PX)** or **(VX)** to ψ. Then $\models \psi' \leftrightarrow \bigvee_{\substack{\alpha \in SD_p \\ \alpha \models \psi}} \alpha'$ where the α' are obtained from the α in the same way as ψ' is obtained from ψ.
Hence it's enough to prove the lemma for **(SN)**, **(PX)** and **(VX)** only for state descriptions.

1. Let $R \in L_d$ and $\alpha \in SD_p$. Let α' be the state description obtained from α via the substitution as in **(SN)** for the relation symbol R.
   For all relations symbols P different from R it holds for all $\vec{a} = a_{i_1}, \ldots a_{i_{arity(P)}}$ with $1 \leq i_1, \ldots, i_{arity(P)} \leq p$ that

   $$\alpha \models P\vec{a} \Longleftrightarrow \alpha' \models P\vec{a}.$$

   Let's see what happens for the relation symbol R. Here for all $\vec{a}, \vec{a}'$

   $$\alpha \models R\vec{a} \leftrightarrow R\vec{a}' \iff \alpha' \models \neg R\vec{a} \leftrightarrow \neg R\vec{a}' \iff \alpha' \models R\vec{a} \leftrightarrow R\vec{a}'.$$

   So overall $S(\alpha) = S(\alpha')$ by for instance the second equivalent definition of the spectral equivalence relation. Hence $w(\alpha) = w(\alpha')$ by **(SX)**.

2. Consider a state description $\mathcal{H} \in SD_p^{01}$ and a state description $\mathcal{H}'$ which is obtained from $\mathcal{H}$ via **(PX)**.
   Then $\mathcal{H}'$ consists of the same arrays as $\mathcal{H}$. The only change is the order of the arrays. But since the spectral equivalence relation S does not depend on the order we have that $S(\mathcal{H}) = S(\mathcal{H}')$. Then clearly the spectra of $\mathcal{H}$ and $\mathcal{H}'$ are the same.

3. Let $\alpha \in SD_p$ and denote by α' the result of a substitution as in **(VX)** and suppose that $R \in L_d$ is the relation symbol for which we do the substitution.
   α and α' decide every literal that does not contain R in the same way.
   Let β, β' be the conjunctions of all literals in α, α' containing R. Suppose that k and j are different over $S(\beta)$ then there exist $1 \leq i_1, \ldots, i_{l-1}, i_{l+1}, \ldots, i_d \leq p$

and different literals λ_1, λ_2 with $\beta \vDash \lambda_1 \wedge \lambda_2$ and such that
$\lambda_1 = Ra_{i_1}, \ldots, a_{i_{l-1}}, a_k, a_{i_{l+1}}, \ldots, a_{i_d}$ and such that
$\lambda_2 = \neg Ra_{i_1}, \ldots, a_{i_{l-1}}, a_j, a_{i_{l+1}}, \ldots, a_{i_d}$ say. Then $\beta' \vDash \sigma(\lambda_1) \wedge \sigma(\lambda_2)$ where

$$\begin{aligned}\sigma(\lambda_1) &= Ra_{i_{\sigma(1)}}, \ldots, a_{i_{\sigma(l)-1}}, a_k, a_{i_{\sigma(l)+1}}, \ldots, a_{i_{\sigma(d)}}\\ \sigma(\lambda_2) &= \neg Ra_{i_{\sigma(1)}}, \ldots, a_{i_{\sigma(l)-1}}, a_j, a_{i_{\sigma(l)+1}}, \ldots, a_{i_{\sigma(d)}} \quad .\end{aligned}$$

Hence a_k and a_j can be distinguished over β'.
The proof that distinguishable a_k, a_j over β' are also distinguishable over β is done in the same way. So $S(\beta) = S(\beta')$ and hence $S(\alpha) = S(\alpha')$.

4. We stated **(AX)** only for state descriptions so we can assume directly that we only consider state descriptions. Let $d, \sigma, \mathcal{H}, \mathcal{G}$ as in **(AX)**. We will show that $S(H_d) = S(G_d)$.
   i, k are in different classes in $S(H_d)$ if and only if there exist tuples $\vec{l}, \vec{l'}$ such that $(H_d)_{\vec{l},i,\vec{l'}} \neq (H_d)_{\vec{l},k,\vec{l'}}$. Since all permutations are bijective this is the case if and only if $\sigma((H_d)_{\vec{l},i,\vec{l'}}) \neq \sigma((H_d)_{\vec{l},k,\vec{l'}})$ equivalently $(G_d)_{\vec{l},i,\vec{l'}} \neq (G_d)_{\vec{l},k,\vec{l'}}$. This however is the definition that i, k are distinguishable over G_d.

5. Consider an arbitrary $\beta \in SD_p$ and put $\beta' := \beta(a_i \mapsto a_q, a_q \mapsto a_i)$ for $i, q \leq p$. Let $\sigma \in S^p$ be the transposition of i and q. Suppose β decides a literal λ then $\beta \models \lambda \iff \beta' \models \lambda(a_i \mapsto a_q, a_q \mapsto a_i)$. So $1 \leq r, s \leq p$ are equivalent over $S(\beta)$ if and only $\sigma(r)$ and $\sigma(s)$ are equivalent over $S(\beta')$.
   But this means that the *sizes* of classes of $S(\beta)$ equal those of the classes of $S(\beta')$. Hence they have the same spectrum and by **(SX)** $w(\beta) = w(\beta')$.
   For an arbitrary $\psi \in QFSL$ with a_q not appearing ψ let $\psi' := \psi(a_i \mapsto a_q)$. Put $N := \max\{const(\psi), q\}$.
   Since $\vDash \psi \leftrightarrow \bigvee_{\substack{\alpha \in SD_N\\ \alpha \vDash \psi}} \alpha$ we have $\vDash \psi' \leftrightarrow \bigvee_{\substack{\alpha \in SD_N\\ \alpha \vDash \psi}} \alpha(a_i \mapsto a_q, a_q \mapsto a_i)$.
   However we have seen above that $w(\alpha) = w(\alpha(a_i \mapsto a_q, a_q \mapsto a_i))$. Hence

$$\begin{aligned}w(\psi') = w(\bigvee_{\substack{\alpha \in SD_N\\ \alpha \vDash \psi}} \alpha(a_i \mapsto a_q, a_q \mapsto a_i)) &= \sum_{\substack{\alpha \in SD_N\\ \alpha \vDash \psi}} w(\alpha(a_i \mapsto a_q, a_q \mapsto a_i))\\ &= \sum_{\substack{\alpha \in SD_N\\ \alpha \vDash \psi}} w(\alpha) = w(\psi) \quad .\end{aligned}$$

□

It follows from the last proof that is enough to prove for all $1 \leq i, k \leq p$ and all $\alpha \in SD_p$ that $w(\alpha) = w(\alpha(a_i \mapsto a_k, a_k \mapsto a_i))$ to show that w satisfies **(CX)**.

So **(SX)** implies all the so far introduced principles with the exception of **(REG)**.

Chapter 4

The Canonical Representation

The idea of a canonical representation was first developed in [17] to prove that **(SX)** implies a *conformity* principle. For a stronger result see Chapter 14.

We here assume that all here considered conjunctions of literals are satisfiable. Before proceeding we recall definition 17 on page 21.

Definition 26. Let $E \in ER_p^t$. Then put $v_s := \min\{r | r \in E_s\}$ for $1 \leq s \leq t$. We define the *very important elements of* E to be the set of these elements $V(E) := \{v_1, v_2, \dots, v_t\}$.
We use $v(i)$ to denote the very important element that is in the same class as i. We call $v(i)$ the *representative of* i. For a tuple of constants $\vec{a} := a_{i_1}, \dots, a_{i_d}$ with $1 \leq i_1, \dots, i_d \leq p$ we write $v(\vec{a})$ for $a_{v(i_1)}, \dots, a_{v(i_d)}$.
Define a function $vip : \{1, \dots, p\} \mapsto \{1, \dots, t\}$ via $vip(i) := n$ if and only if $i \in E_n$, equivalently $v(i) = v_n$. Again we put for a tuple of constant $\vec{a}$ as above $vip(\vec{a}) := a_{vip(i_1)}, \dots, a_{vip(i_d)}$. Note for later use that

$$vip(v(i)) = vip(i) \ .$$

If there can be some confusion we'll write v_E and vip_E to make clear that we work with respect to an equivalence relation E.

Since we assume that classes of an equivalence relation are ordered in a certain way we have for all equivalence relations that $v_n \geq n \geq v(n),\ vip(1) = v_1 = 1$.

Definition 27. Let $\psi \in QFSL$ be a conjunction of literals. Let $\psi \upharpoonright_{i_1,\dots,i_k}$ be the conjunction of those literals of ψ that only contain constants in $\{a_{i_1}, \dots, a_{i_k}\}$.
If $\{i_1, \dots, i_k\} = \{1, \dots, p\}$ for some $p \in \mathbb{N}$ we also write $\psi \upharpoonright_p$.

Definition 28. Let $\alpha \in QFSL$ be a conjunction of literals with $const(\alpha) \leq p$. Then define $\frac{\alpha}{E} := \alpha \upharpoonright_{V(E)}$ for $E \in ER_p$.

Definition 29. Let $\alpha \in QFSL$ be a conjunction of literals with $const(\alpha) \leq p$ and let $E \in ER_p$. Then define α/E to be the conjunction of literals satisfying for all d and all $d-$ary relations R and all $1 \leq i_1, \ldots, i_d \leq |E|$

$$\alpha/E \models Ra_{i_1} \ldots a_{i_d} \iff \alpha \models Ra_{v_{i_1}} \ldots a_{v_{i_d}} \tag{4.1}$$
$$\alpha/E \models \neg Ra_{i_1} \ldots a_{i_d} \iff \alpha \models \neg Ra_{v_{i_1}} \ldots a_{v_{i_d}} \tag{4.2}$$

and if α doesn't decide $Ra_{v_{i_1}} \ldots a_{v_{i_d}}$ then α/E doesn't decide $Ra_{i_1} \ldots a_{i_d}$. We say that α is *divided* by E.
If $\alpha \in SD_p(L)$ then α/E decides $Ra_{j_1} \ldots a_{j_{arity(R)}}$ if and only if $\{j_1, \ldots, j_{arity(R)}\} \subset \{1, \ldots, |E|\}$ and hence $\alpha/E \in SD_{|E|}(L)$.

Assume that in our last definitions α is a state description on p. Then in Definition 29 we take $\alpha \upharpoonright_{V(E)}$ and make it into a state description by renaming constants. From now we will mainly work with $/E$ since it is a state description in that case.
For $E \in ER_p^t$ and $\alpha \in SD_p$ we have $\frac{\alpha}{E}(a_{v_1} \mapsto a_1, \ldots, a_{v_t} \mapsto a_t) = \alpha/E$.
To illustrate these rather technical and not very transparent definitions we give the following

Example 6. We consider $H \in SD_4(L_3)$ where the first array is $(H)_{1,i_2,i_3}$ the second is $(H)_{2,i_2,i_3}$ and the third $(H)_{3,i_2,i_3}$ and finally the fourth $(H)_{4,i_2,i_3}$

$$\begin{array}{cccc} 1 & 6 & 6 & 1 \\ 2 & 5 & 5 & 5 \\ 2 & 5 & 5 & 5 \\ 1 & 1 & 1 & 1 \end{array} \qquad \begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{cccc} 4 & 5 & 5 & 5 \\ 2 & 5 & 5 & 5 \\ 2 & 5 & 5 & 5 \\ 0 & 8 & 8 & 7 \end{array}$$

Then $S(H) = \langle\{1\}, \{2,3\}, \{4\}\rangle$, $v_1 = 1$, $v_2 = 2$, $v_3 = 4$ and hence $\frac{H}{S(H)}$ is

$$\begin{array}{cccc} 1 & 6 & & 1 \\ 2 & 5 & & 5 \\ & & & \\ 1 & 1 & & 1 \end{array} \qquad \begin{array}{cccc} 0 & 0 & & 1 \\ 0 & 0 & & 0 \\ & & & \\ 0 & 0 & & 0 \end{array} \qquad \begin{array}{cccc} & & & \\ & & & \\ & & & \\ & & & \end{array} \qquad \begin{array}{cccc} 4 & 5 & & 5 \\ 2 & 5 & & 5 \\ & & & \\ 0 & 8 & & 7 \end{array}$$

where a blank at position i_1, i_2, i_3 means that the atom on $a_{i_1}, a_{i_2}, a_{i_3}$ has been deleted. Here the third block is missing completely since 2 and 3 are equivalent over $S(H)$.

$H/S(H)$ is

$$\begin{matrix} 1 & 6 & 1 \\ 2 & 5 & 5 \\ 1 & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 4 & 5 & 5 \\ 2 & 5 & 5 \\ 0 & 8 & 7 \end{matrix}$$

For $E \in ER^4_4$ with $E = \langle\{1\}, \{2\}, \{3\}, \{4\}\rangle$ we have $H/E = H$.
Writing the example above in first order logic would take considerably more space without providing new insights.

Lemma 7. For all $\alpha \in SD^t_p(L)$ we have that $\alpha/S(\alpha) \in SD^t_t(L)$.

Proof. We already know that $\alpha/S(\alpha) \in SD_t$ and hence have to show that the spectrum of $\alpha/S(\alpha)$ is $1@t$. In this lemma v and vip are always with respect to the equivalence relation $S(\alpha)$.
Therefore suppose that a_i and a_k are not equivalent over α. Hence there are a relation symbol $R \in L$ and tuples $\vec{a}, \vec{a}'$ such that $\alpha \models R\vec{a}a_i\vec{a}' \wedge \neg R\vec{a}a_k\vec{a}'$ say.
Then $\alpha \models Rv(\vec{a})a_iv(\vec{a}') \wedge \neg Rv(\vec{a})a_kv(\vec{a}')$ and hence
$\alpha \models Rv(\vec{a})v(a_i)v(\vec{a}') \wedge \neg Rv(\vec{a})v(a_k)v(\vec{a}')$.
Since a_i and a_k are distinguishable over α it holds that $v(a_i) \neq v(a_k)$ and hence $vip(a_i) \neq vip(a_k)$ which in turn implies

$$\alpha/S(\alpha) \models Rvip(\vec{a})vip(a_i)vip(\vec{a}') \wedge \neg Rvip(\vec{a})vip(a_k)vip(\vec{a}') \ .$$

So $vip(a_i)$ is distinguishable from $vip(a_k)$ over $\alpha/S(\alpha)$. Hence whenever two constants a_i, a_k look different over α so do $a_{vip(i)}, a_{vip(k)}$ over $\alpha/S(\alpha)$.
Since $|S(\alpha)| = t$ we find that $S(\alpha/S(\alpha))$ has t classes. □

Definition 30. We now define a map $Red : SD^t_p(L_d) \rightarrow SD^t_t(L_d) \times ER^t_p$ such that $H \mapsto \langle H/S(H), S(H)\rangle$. We call the map $Red(H)$ the *reduction of* H. We call $\langle H/S(H), S(H)\rangle$ the *canonical representation* of H. $H/S(H)$ is referred to as the *canonical subarray* of H.[1]

Definition 31. For $q \leq p$ we define a map $BlowUp : SD_q(L_d) \times ER^q_p \rightarrow SD_p(L_d)$ by the following construction. We denote the *blow up* of $\gamma \in SD_q(L_d)$ by $E \in ER^q_p$ by $\gamma \cdot E$. We define $\gamma \cdot E$ to be the unique state description in $SD_p(L_d)$ that for all $R \in L_d$

[1] or canonical subformula if we want to work with first order formulae.

and all $1 \leq i_1, \ldots, i_d \leq p$ satisfies that

$$\gamma \cdot E \vDash Ra_{i_1} \ldots a_{i_d} \text{ if and only if } \gamma \vDash Ra_{vip_E(i_1)} \ldots a_{vip_E(i_d)} \ . \tag{4.3}$$

An example can be found after the next proof.

Lemma 8. The map Red is bijective and for $\alpha \in SD_p(L_d)$ we have

$$BlowUp(Red(\alpha)) = (\alpha/S(\alpha)) \cdot S(\alpha) = \alpha \ . \tag{4.4}$$

Proof. First we check that the function that Red is injective.
Therefore consider different $\alpha, \beta \in SD_p(L_d)$ such that $S(\alpha) \neq S(\beta)$. Then clearly $Red(\alpha) \neq Red(\beta)$.
So let's assume $S(\alpha) = S(\beta)$. Since $\alpha \neq \beta$ there are a relation symbol $R \in L$ and a tuple $\vec{a} := a_{i_1}, \ldots, a_{i_d}$ such that $\alpha \models R\vec{a}$ and $\beta \models \neg R\vec{a}$ say. Then $\alpha \models Rv_{S(\alpha)}(\vec{a})$ and $\beta \models \neg Rv_{S(\beta)}(\vec{a})$, so $\beta \models \neg Rv_{S(\alpha)}(\vec{a})$.
Since as noted above $vip(v(\vec{a})) = vip(\vec{a})$ we have that $\alpha/S(\alpha) \models Rvip_{S(\alpha)}(\vec{a})$ and $\beta/S(\alpha) \models \neg R\vec{v}ip_{S(\alpha)}(a)$. Hence $\alpha/S(\alpha) \neq \beta/S(\alpha) = \beta/S(\beta)$.
Next we show that Red is surjective. We'll therefore prove that for all $1 \leq t \leq p$ the map $BlowUp : SD_t^t(L_d) \times ER_p^t \to SD_p^t(L_d)$ is injective and that for $\alpha \in SD_p(L_d)$ we have $(\alpha/S(\alpha)) \cdot S(\alpha) = \alpha$ and that for $\beta \in SD_t^t$ and $E \in ER_p^t$ $(\beta \cdot E)/E = \beta$ holds.
So consider a $\gamma \in SD_t^t(L_d)$ and an $E \in ER_p^t$. From now on we always work with respect to E until stated otherwise and subsequently we drop the index E.
Now consider $1 \leq i, k \leq p$ that are equivalent over E, which implies $v(i) = v(k)$. Then for every relation symbol $R \in L_d$ and all tuples $\vec{a}, \vec{a}'$ we have by the definition of $\gamma \cdot E$ that $\gamma \models Rvip(\vec{a})vip(a_i)vip(\vec{a}')$ if and only if $\gamma \cdot E \models R\vec{a}a_i\vec{a}'$ if and only if $\gamma \cdot E \models R\vec{a}a_k\vec{a}'$.
Hence i and k are equivalent over $S(\gamma \cdot E)$.
Now assume that $1 \leq i, k \leq p$ are in different classes of E and we have hence $v(i) \neq v(k)$. Since $\gamma \in SD_t^t(L_d)$ there are tuples $\vec{a}, \vec{a}'$ and a relation symbol R in L_d such that say $\gamma \models Rvip(\vec{a})a_{vip(i)}vip(\vec{a}') \wedge \neg Rvip(\vec{a})a_{vip(k)}vip(\vec{a}')$.
But then by definition of the blow up $\gamma \cdot E \models R\vec{a}a_i\vec{a}' \wedge \neg R\vec{a}a_k\vec{a}'$. So a_i and a_k are not equivalent over $\gamma \cdot E$. So overall $S(\gamma \cdot E) = E$, and so $\gamma \cdot E \in SD_p^t(L_d)$ as claimed.
Next we want to show that the blow up is injective. We have to show that for different $\gamma, \delta \in SD_t^t$ and $E \in ER_p^t$ that $\gamma \cdot E \neq \delta \cdot E$ since we already know that $S(\gamma \cdot E) = E$ and so for different $F, F' \in ER_p^t$ we have $\gamma \cdot F \neq \gamma \cdot F'$ and $\gamma \cdot F \neq \delta \cdot F'$.

There are a relation symbol R and $a_{i_1}, \dots a_{i_d}$ with $1 \le i_1, \dots, i_d \le p$ such that $\gamma \models Rvip(\vec{a})$ and $\delta \models \neg Rvip(\vec{a})$ say. Then $\gamma \cdot E \models R\vec{a}$ and $\delta \cdot E \models \neg R\vec{a}$ implying $\gamma \cdot E \neq \delta \cdot E$.
Hence for $\alpha \in SD_p^t(L_d)$ and all relation symbols $R \in L_d$ and all tuples $\vec{a}$ we have

$$\alpha \models R\vec{a} \iff \alpha/S(\alpha) \models Rvip_{S(\alpha)}(\vec{a}) \iff (\alpha/S(\alpha)) \cdot S(\alpha) \models R\vec{a} \ .$$

For $\beta \in SD_t^t$ and $E \in ER_p^t$ $(\beta \cdot E)/E = \beta$ follows directly from the definitions. $\square$

Example 7. Suppose L_1 contains at least four unary relation symbols and let $G \in SD_5^5(L_1)$ such that $G = \langle 2, 1, 0, 5, 9 \rangle$ and $E \in ER_8^5$ such that $2, 4$, and 7 are equivalent and so are 6 and 8. Then $G \cdot E = \langle 2, 1, 0, 1, 5, 9, 1, 9 \rangle$.
Suppose $J \in SD_3^3(L_3)$ and $F \in ER_5^3$ such that $F = \langle \{1, 2\}, \{3, 5\}, \{4\} \rangle$ and

$$J = \begin{array}{ccc} 1 & 6 & 1 \\ 2 & 5 & 5 \\ 1 & 1 & 1 \end{array} \qquad \begin{array}{ccc} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \qquad \begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array}$$

then $J \cdot F$ equals

$$\begin{array}{ccccc} 1 & 1 & 6 & 1 & 6 \\ 1 & 1 & 6 & 1 & 6 \\ 2 & 2 & 5 & 5 & 5 \\ 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 5 & 5 & 5 \end{array} \qquad \begin{array}{ccccc} 1 & 1 & 6 & 1 & 6 \\ 1 & 1 & 6 & 1 & 6 \\ 2 & 2 & 5 & 5 & 5 \\ 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 5 & 5 & 5 \end{array} \qquad \begin{array}{ccccc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

$$\begin{array}{ccccc} 1 & 1 & 2 & 3 & 2 \\ 1 & 1 & 2 & 3 & 2 \\ 4 & 4 & 5 & 6 & 5 \\ 7 & 7 & 8 & 9 & 8 \\ 4 & 4 & 5 & 6 & 5 \end{array} \qquad \begin{array}{ccccc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

where the three blocks in the first row are $(J \cdot F)_{1ij}$, $(J \cdot F)_{2ij}$ and $(J \cdot F)_{3ij}$ respectively with $1 \le i, j \le 5$. $\frac{J \cdot F}{F}$ equals

$$\begin{array}{ccccc} 1 & @ & 6 & 1 & @ \\ @ & @ & @ & @ & @ \\ 2 & @ & 5 & 5 & @ \\ 1 & @ & 1 & 1 & @ \\ @ & @ & @ & @ & @ \end{array} \qquad \begin{array}{ccccc} @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \end{array} \qquad \begin{array}{ccccc} 0 & @ & 0 & 1 & @ \\ @ & @ & @ & @ & @ \\ 0 & @ & 0 & 0 & @ \\ 0 & @ & 0 & 0 & @ \\ @ & @ & @ & @ & @ \end{array}$$

$$\begin{array}{ccccc} 1 & @ & 2 & 3 & @ \\ @ & @ & @ & @ & @ \\ 4 & @ & 5 & 6 & @ \\ 7 & @ & 8 & 9 & @ \\ @ & @ & @ & @ & @ \end{array} \qquad \begin{array}{ccccc} @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \\ @ & @ & @ & @ & @ \end{array}$$

where we used the $@-$symbol to help the visualization of the blank spots. So we deleted in $J \cdot F$ every literal that contains a_2 or a_5 as two and five have a representative that is smaller than two respectively five.

We now go ahead and extend the ideas above to a general polyadic language.

Definition 32. We let $Red : SD_p^t(L) \to SD_t^t(L) \times ER_p^t$ be the map defined by $\mathcal{H} \mapsto \langle \mathcal{H}/S(\mathcal{H}), S(\mathcal{H}) \rangle$. We call $\langle \mathcal{H}/S(\mathcal{H}), S(\mathcal{H}) \rangle$ the *canonical representation* of $\mathcal{H}$ and $Red(\mathcal{H})$ the *reduction of* $\mathcal{H}$. $\mathcal{H}/S(\mathcal{H})$ is referred to as the *canonical subarray* of $\mathcal{H}$.[2]

Definition 33. Let $\alpha \in QFSL$ be a conjunction of literals with $const(\alpha) = p$ and let $E \in ER_q$ with $q \geq p$. Then let the *blow up of* α *with* E, denoted by $\alpha \cdot E$, be the conjunction of literals with $const(\alpha \cdot E) \leq q$ satisfying for all relation symbols $R \in L$ and all $\vec{a}$ that

$$\alpha \cdot E \models R\vec{a} \iff \alpha \models Rvip_E(\vec{a}) \tag{4.5}$$

$$\alpha \cdot E \models \neg R\vec{a} \iff \alpha \models \neg Rvip_E(\vec{a}) \tag{4.6}$$

with the convention that if $Rvip_E(\vec{a})$ is not decided by α then $\alpha \cdot E$ does not decide $R\vec{a}$.
$R\vec{a}$ is for instance not decided by $\alpha \cdot E$ if there is an a_i in $\vec{a}$ such that $vip_E(i) > p$.

Note that if $\alpha \in SD_p(L)$ and $E \in ER_q^p$, then $\alpha \cdot E \in SD_q(L)$.

[2]or canonical subformula if we want work with first order formulae.

Theorem 2. The Canonical Representation
The map $Red : SD_p^t(L) \to SD_t^t(L) \times ER_p^t$ *with* $Red(\alpha) = \langle \alpha/S(\alpha), S(\alpha)\rangle$ *is bijective and for* $\alpha \in SD_p(L)$ *we have* $(\alpha/S(\alpha)) \cdot S(\alpha) = \alpha$.

Proof. The proof is almost the same as the one for the lemma above. Whenever we quantified over all $R \in L_d$ in the above proof we here have to quantify over all $R \in L$, and whenever we stated that there was one $R \in L_d$ we here have that R is in L.
When we considered tuples of constants of length d or $d-1$ we here have to consider tuples of constants of length $arity(R)$ respectively $arity(R) - 1$.
These two small changes don't make any difference to the proof itself and we can now conclude by copying the proof from above with these two adaptations. □

Corollary 1. For all equivalence relations $E \in ER_p^t$ with spectrum $\vec{x} \in X_p^t$ we have [3]

$$|\{\alpha \in SD_p(L) \mid S(\alpha) = E\}| = |SD_{|E|}^{|E|}(L)| = N(\emptyset, 1@|E|, L) \quad \text{and} \tag{4.7}$$

$$|\{\alpha \in SD_p^t(L) \mid \overrightarrow{S(\alpha)} = \vec{x}\}| = \binom{p}{\vec{x}} \frac{|SD_t^t(L)|}{\prod_{i=1}^p q_i!} \tag{4.8}$$

where $q_i := |\{j \mid x_j = i\}|$ with the standard convention that $0! = 1$.
(4.8) holds for purely unary L if and only if $t \leq 2^{m_1}$. If L is not purely unary then (4.8) always holds.

Proof. By the theorem above the first equation follows directly and for the second we find

$$\begin{aligned} |\{\alpha \in SD_p^t \mid \overrightarrow{S(\alpha)} = \vec{x}\}| &= \sum_{\substack{E \in ER_p^t \\ \overrightarrow{S(E)} = \vec{x}}} |\{\alpha \in SD_p^t \mid S(\alpha) = E\}| \\ &= \sum_{\substack{E \in ER_p^t \\ \overrightarrow{S(E)} = \vec{x}}} |SD_t^t| = \binom{p}{\vec{x}} \frac{|SD_t^t|}{\prod_{i=1}^p q_i!} \quad . \end{aligned}$$

Where for the last step we count how many equivalence relations have spectrum $\vec{x}$.
Therefore we first have to chose for every element how big the class is it will belong to; yielding the multinomial factor.
Then we have to correct for classes of the same size, that's why we have to divide by the $q_i!$.
If L is purely unary then $SD^{>2^{m_1(L)}}$ is empty. □

[3] For a definition of $N(\emptyset, 1@|E|, L)$ see definition 35 on page 46.

Lemma 9. Let $\alpha \in SD_p$ and let $E \in ER_q$ with $q \geq p$ such that $S(\alpha) \leq E \restriction_p$. Then $(\alpha/E) \cdot E \vDash \alpha$.

Proof. Since we assumed $S(\alpha) \leq E \restriction_p$ we find for $\vec{a} = a_{i_1}, \ldots, a_{i_d}$ with $1 \leq i_1, \ldots, i_d \leq p$ that

$$\begin{aligned} \alpha \models R\vec{a} &\iff \alpha \models Rv_{S(\alpha)}(\vec{a}) &(4.9)\\ &\iff \alpha \models Rv_E(\vec{a}) &(4.10)\\ &\iff \alpha/E \models Rvip_E(\vec{a}) &(4.11)\\ &\iff (\alpha/E)\cdot E \models R\vec{a} &(4.12)\end{aligned}$$

where we used $S(\alpha) \leq E \restriction_p$ to obtain (4.10), $1 \leq i, k \leq p$ that are equivalent over $E \restriction_p$ are equivalent over $S(\alpha)$. □

Remark 5. Suppose in the lemma above we use an equivalence relation F on p that is not a refinement of $S(\alpha)$. Suppose for example $1, 2 \in F_1$ but we can tell a_1 and a_2 over α apart.

Then there are two tuples of constants $\vec{a}, \vec{a}'$ and a relation symbol P such that say $\alpha \models P\vec{a}a_1\vec{a}' \wedge \neg P\vec{a}a_2\vec{a}'$. So we find $\alpha/F \models Pvip_F(\vec{a})a_1vip_F(\vec{a}')$ and hence $(\alpha/F) \cdot F \models P\vec{a}a_2\vec{a}'$ since $vip_F(2) = 1$.

So the blow up is not the inverse of the reduction. That's why F is in most applications a refinement of $S(\alpha)$.

Lemma 10. Let $\alpha \in SD_p$ and let $E \in ER_q^p$. Then $S(\alpha \cdot E) \leq E$.

Proof. Suppose $1 \leq i, j \leq q$ are equivalent over E and hence $vip_E(i) = vip_E(j)$. Then for all relation symbols R and all tuples $\vec{a}, \vec{a}'$ we have

$$\alpha \models Rvip_E(\vec{a})vip_E(a_i)vip_E(\vec{a}') \iff \alpha \cdot E \models R\vec{a}a_i\vec{a}' \iff \alpha \cdot E \models R\vec{a}a_j\vec{a}' \ .$$

Hence i and j are equivalent over $S(\alpha \cdot E)$ and so $S(\alpha \cdot E) \leq E$. □

Lemma 11. For $E \in ER_p$ the following map f is a bijection. $f : \{\alpha \in SD_p | S(\alpha) \leq E\} \to SD_{|E|}$ with $f(\alpha) := \alpha/E$. Hence $|\{\alpha \in SD_p | S(\alpha) \leq E\}| = |SD_{|E|}|$.

Proof. Consider different $\alpha, \alpha' \in SD_p$ such that $S(\alpha) \leq E$ and $S(\alpha') \leq E$. Then there exist a tuple $\vec{a}$ and a relation symbol R such that $\alpha \vDash R\vec{a}$ and $\alpha' \vDash \neg R\vec{a}$ say. Then $\alpha \vDash Rv_E(\vec{a})$ and $\alpha' \vDash \neg Rv_E(\vec{a})$.

Hence $\alpha/E \vDash Rvip_E(\vec{a})$ and $\alpha'/E \vDash \neg Rvip_E(\vec{a})$. This implies that f is injective.

Now consider a $\beta \in SD_{|E|}$. Then $\beta \cdot E \in \{\alpha \in SD_p | S(\alpha) \leq E\}$ by Lemma 10. But then by our definitions of blow up and reduction $(\beta \cdot E)/E = \beta$. Hence f is surjective. $\square$

Chapter 5

A first set of Representation Theorems

The following representation theorems were proved in [22] for purely binary languages. We here drop the restriction on the arity of the relation symbols in L.

5.1 Introducing $N(\vec{x}, \vec{y}, L)$

The next step of the way will be showing that

$$|\{\beta \in SD_q(L) \mid \beta \vDash \alpha, \overrightarrow{\mathcal{S}(\beta)} = \vec{y}\}|$$

only depends on the spectrum $\vec{x} \in X_p$ of $\alpha \in SD_p(L)$. We will define $N(\vec{x}, \vec{y}, L)$ to be that number.

Lemma 12. Let $H \in SD_p^t(L_d)$. Then there exists no $H' \in SD_{p+1}^{\geq t+1}(L_d)$ satisfying

1. $H = H' \restriction_p$,
2. $S(H')(p+1)\ n$ for some $1 \leq n \leq p$ and
3. $S(H) < S(H') \restriction_p$.

Furthermore the same holds for $\mathcal{H} \in SD_p^t(L)$ in place of H and $\mathcal{H}' \in SD_{p+1}^{\geq t+1}(L)$ instead of $H' \in SD(L_d)$.

Proof. That is clear for $d = 1$. If L is purely unary then the new constant can never split an equivalence class .
Suppose from now on that $d \geq 2$.
We do a proof by contradiction, so let's assume there is such an array H'.

Since $S(H')\restriction_p$ is a proper refinement of $S(H)$ at least one class of $S(H)$ has to get split, say $S(H)ij$ but not $S(H')\restriction_p ij$ with $1 \le i < j \le p$. Then w.l.o.g. there are tuples $\vec{x}, \vec{y}$ of combined length $d-2$ such that $(H')_{\vec{x},i,\vec{y},p+1} \neq (H')_{\vec{x},j,\vec{y},p+1}$.[1] From $S(H')(p+1)n$ it follows that

$$(H')_{\vec{x},i,\vec{y},n} = (H')_{\vec{x},i,\vec{y},p+1} \neq (H')_{\vec{x},j,\vec{y},p+1} = (H')_{\vec{x},j,\vec{y},n} \quad . \tag{5.1}$$

Now we replace every occurrence of $p+1$ in $\vec{x}$ and $\vec{y}$ by n. Under this substitution $\vec{x}, \vec{y}$ transform to $\vec{x}', \vec{y}'$. But then since $S(H')(p+1)n$

$$(H')_{\vec{x}',i,\vec{y}',n} = (H')_{\vec{x},i,\vec{y},n} = (H')_{\vec{x},i,\vec{y},p+1} \neq (H')_{\vec{x},j,\vec{y},p+1} = (H')_{\vec{x},j,\vec{y},n} = (H')_{\vec{x}',j,\vec{y}',n}$$

where $(H')_{\vec{x}',i,\vec{y}',n}$ and $(H')_{\vec{x}',j,\vec{y}',n}$ are entries in H. So $(H)_{\vec{x}',i,\vec{y}',n} \neq (H)_{\vec{x}',j,\vec{y}',n}$.
If $\vec{x}, \vec{y}$ didn't contain any $p+1$ we can skip the substitution and conclude directly that $(H')_{\vec{x},i,\vec{y},n}$ and $(H')_{\vec{x},j,\vec{y},n}$ are in H.
Since we assumed that i and j are in the same equivalence class over $S(H)$ and $(H)_{\vec{x}',i,\vec{y}',n} \neq (H)_{\vec{x}',j,\vec{y}',n}$ we have a contradiction.
For the second part of the proof for $\mathcal{H}, \mathcal{H}'$ observe that $p+1$ has to be equivalent to the same n over every non-empty array of $\mathcal{H}$. The split of one class has to happen over at least one such array. But we have just seen that this can't happen. □

There's no mathematical need to give the next lemma. However we feel that it provides the reader with a warm-up that prepares the ground for the general case. In the general case we will be interested in state descriptions over L and not just L_d.

Lemma 13. Let $H, H' \in SD_p(L_d)$ such that $S(H) = S(H')$. Let $E \in ER_q$ where $q > p$, and let $C_d(H,E) := \{J \in SD_q(L_d) \mid S(J) = E \text{ and } J\restriction_p = H\}$.[2] Then

$$|C_d(H,E)| = |C_d(H',E)| \quad . \tag{5.2}$$

Proof. We can assume $q = p+1$ since a proof for $q = p+1$ ensures the result for all $q > p$ since we can inductively do this proof over and over again.
First assume that $d = 1$ and assume that there is an $i \le p$ such that $Ei(p+1)$. Then for all $R \in L_1$ $J \in C_1(H,E)$ has to decide Ra_{p+1} the same way H decides Ra_i.[3] Hence

[1] The w.l.o.g. refers to the appearance of $p+1$ as the last element of the tuple. That's assumed for notational convenience only.
[2] We of course assume that $L_d \neq \emptyset$.
[3] In array formulation we would write $(J)_{p+1}$ has to equal $(H)_i$.

$C_1(H,E) = 1 = C_1(H',E)$.
If $p+1$ is a singleton over E then we have a free choice among the not yet chosen atoms. And since $S(H) = S(H')$ implying that H and H' have the same number of different entries, $|C_1(H,E)| = 2^{m_1} - |S(H)| = |C_1(H',E)|$ follows.
From now on assume $d \geq 2$.
The proof is done by induction on the number of equivalence classes of E, where we can assume that E has at least as many classes as $S(H)$. We can assume that $S(H) \leq E \upharpoonright_p$ since every two constants that can be distinguished over H can be told apart over every extension.
For the **base case** we have $|E| = |S(H)|$. Then $E \upharpoonright_p = S(H) = S(H')$ and hence $|C_d(H,E)| = |C_d(H',E)| = 1$ since $p+1$ has to join an existing equivalence class and for a fixed array H there is only one such extension J.
Now for **the induction step** let $E \in ER_{p+1}$ such that $|E| > |S(H)| = |S(H')|$.
Due to Lemma 12 we can assume that $p+1 \in V(E)$, that is $p+1$ does not join any class.
Next we show:

$$\sum_{\substack{F \in ER_{p+1} \\ F \leq E}} |C_d(H,F)| = |\{G \in SD_{|E|}(L_d) \mid H \subset G \cdot E\}| \tag{5.3}$$

$$= \frac{|SD_{|E|}(L_d)|}{|SD_{|E|-1}(L_d)|} = (2^{m_d})^{|E|^d - (|E|-1)^d} \tag{5.4}$$

$$= |\{G \in SD_{|E|}(L_d) \mid H' \subset G \cdot E\}| \tag{5.5}$$

$$= \sum_{\substack{F \in ER_{p+1} \\ F \leq E}} |C_d(H',F)| \quad . \tag{5.6}$$

The hard work we invested in the canonical representation section will now start paying dividends. The first and the last equality follow from Lemma 11 on page 39.
There are $(2^{m_d})^{(|E|^d)}$ arrays G in $SD_{|E|}(L_d)$. For their blow up with E it holds that $S(G \cdot E) \leq E$ by Lemma 10 on page 39.
But not all $G \cdot E$ contain H or H'. However if $H/E \subset G$ then $H \subset G \cdot E$ by Lemma 9 on page 39.
If H/E is not a subarray of G then there is a tuple $1 \leq i_1, \ldots, i_d \leq p$ such that $(H/E)_{vip_E(i_1),\ldots,vip_E(i_d)} \neq (G)_{vip_E(i_1),\ldots,vip_E(i_d)}$. So $(H)_{i_1,\ldots,i_d} = ((H/E) \cdot E)_{i_1,\ldots,i_d} \neq (G \cdot E)_{i_1,\ldots,i_d}$. So $G \cdot E$ does not contain H. We have hence shown that (and similarly

for H')

$$\{G \in SD_{|E|}(L_d) \mid H \subset G \cdot E\} = \{G \in SD_{|E|}(L_d) \mid H/E \subset G\} \ .$$

H is an array on p and $S(H) \leq E \upharpoonright_p$. So H/E is a $d-$dimensional array on $|E|-1$ since $p+1$ is a singleton over E. Hence H/E contains $(|E|-1)^d$ entries yielding the remaining equations.
The induction hypothesis $|C_d(H,F)| = |C_d(H',F)|$ for all $F \in ER_{p+1}$ with $|F| < |E|$ hence entails $|C_d(H,E)| = |C_d(H',E)|$. □

We will now do the general case. The main ideas are already in the above proof.

Lemma 14. For $p < q$ and an equivalence relation E on q and $\mathcal{H}, \mathcal{H}' \in SD_p$ let $D(\mathcal{H},E) := \{\mathcal{H}'' \in SD_q \mid \mathcal{H}'' \upharpoonright_p = \mathcal{H} \text{ and } S(\mathcal{H}'') = E\}$. If $S(\mathcal{H}) = S(\mathcal{H}')$ then

$$|D(\mathcal{H},E)| = |D(\mathcal{H}',E)| \ . \tag{5.7}$$

Proof. Again it's enough to prove the lemma for $q = p+1$ which we will do by induction on $|E|$. Furthermore we can assume that $S(\mathcal{H}) \leq E \upharpoonright_p$, since every two constants that can be distinguished over $\mathcal{H}$ can be told apart over every extension.
Begin of the induction: Assume that $|E| = |S(\mathcal{H}')|$ then $p+1$ has to join an already existing class and hence $E \upharpoonright_p = S(\mathcal{H}')$. Then the lemma holds since $|D(\mathcal{H},E)| = |D(\mathcal{H}',E)| = 1$, since $p+1$ has to join the same class over every array, and for each array there is only one way of doing so.
For the **induction step** we have $|E| > |S(\mathcal{H}')|$.
Here we can assume that $p+1$ is a singleton over E. We will show that

$$\begin{aligned}\sum_{\substack{F \in ER_{p+1} \\ F \leq E \\ S(\mathcal{H}') \leq F\upharpoonright_p}} |D(\mathcal{H},F)| &= \prod_{d=1}^{r_0} \frac{|SD_{|E|}(L_d)|}{|SD_{|E|-1}(L_d)|} = \prod_{d=1}^{r_0} (2^{m_d})^{|E|^d - (|E|-1)^d} \\ &= \sum_{\substack{F \in ER_{p+1} \\ F \leq E \\ S(\mathcal{H}') \leq F\upharpoonright_p}} |D(\mathcal{H}',F)| \ . \end{aligned} \tag{5.8}$$

We will count the number of elements of $\{\mathcal{J} \in SD_{|E|}(L) | \mathcal{H} \subset \mathcal{J} \cdot E\}$ respectively $\mathcal{H}' \subset \mathcal{J} \cdot E$. By Lemma 11 on page 39 it is enough to show that $|\{\mathcal{J} \in SD_{|E|}(L) | \mathcal{H} \subset \mathcal{J} \cdot E\}| = \frac{|SD_{|E|}(L)|}{|SD_{|E|-1}(L)|}$ in order to prove (5.8).
We have $|SD_{|E|}(L)| = \prod_{d=1}^{r_0} (2^{m_d})^{(|E|^d)}$. So it's enough to ensure that $\mathcal{H} \subset \mathcal{J} \cdot E$

respectively $\mathcal{H}' \subset \mathcal{J} \cdot E$.
Now if $\mathcal{H}/E \subset \mathcal{J}$ then certainly $\mathcal{H} \subset \mathcal{J} \cdot E$.
On the other hand if $\mathcal{H}/E$ is not a subarray of $\mathcal{J}' \in SD_{|E|}$ then there exists a $1 \leq d \leq r_0$ and $1 \leq i_1, \ldots, i_d \leq p$ such that $(H_d/E)_{vip_E(i_1),\ldots,vip_E(i_d)} \neq (J'_d)_{vip_E(i_1),\ldots,vip_E(i_d)}$. Then $(H_d)_{i_1,\ldots,i_d} \neq (J'_d \cdot E)_{i_1,\ldots,i_d}$ and hence $\mathcal{J}' \notin \{\mathcal{J} \in SD_{|E|} | \mathcal{H} \subset \mathcal{J} \cdot E\}$. So

$$\{\mathcal{J} \in SD_{|E|} | \mathcal{H} \subset \mathcal{J} \cdot E\} = \{\mathcal{J} \in SD_{|E|} \mid \mathcal{H}/E \subset \mathcal{J}\} \ . \tag{5.9}$$

$\mathcal{H}/E \in SD_{|E|-1}(L)$ since $\mathcal{H} \in SD_p(L)$ and $p+1$ is a singleton over E. Hence these sets have $\frac{|SD_{|E|}(L)|}{|SD_{|E|-1}(L)}$ many elements. Hence we have proved (5.8).
Now the induction hypothesis for all $F < E$ implies $|D(\mathcal{H}, E)| = |D(\mathcal{H}', E)|$. □

Definition 34. Let $\sigma \in S^p$ and $\alpha \in SD_p$. Then define $\alpha(\sigma)$ to be the result of simultaneously replacing every constant a_i in α by $a_{\sigma(i)}$. Note that $\alpha(\sigma) \in SD_p$ and that $\alpha(a_i \mapsto a_j)$ is consistent if and only if $\alpha(\sigma)(a_{\sigma(i)} \mapsto a_{\sigma(j)})$ is.
Next we want to extend σ to a permutation σ_q in S^q for $q \geq p$. For $1 \leq i \leq p$ simply put $\sigma_q(i) := \sigma(i)$. And for $p+1 \leq i \leq q$ let $\sigma_q(i) := i$.
For $E \in ER_q$ and $\pi \in S^q$ define $\pi(E)$ to be the following equivalence relation on q. $\pi(E)ij$ holds if and only if $E\pi^{-1}(i)\pi^{-1}(j)$ holds.
We hence have $S(\alpha(\sigma)) = \sigma(S(\alpha))$.

Lemma 15. For $p \leq q$, $\vec{x} \in X_q$ let $\alpha, \beta \in SD_p$ such that α and β have the same spectrum. Put $G(\alpha, \vec{x}) := \{\gamma \in SD_q \mid \gamma \vDash \alpha \text{ and } \overrightarrow{\mathcal{S}(\gamma)} = \vec{x}\}$. Then $|G(\alpha, \vec{x})| = |G(\beta, \vec{x})|$.

Proof. Since α and β have the same spectrum there is $\sigma \in S^p$ such that $\sigma(S(\alpha)) = S(\beta)$. Then $S(\alpha(\sigma)) = S(\beta)$.
Note that for $\delta \in SD_q(L)$ $\delta \vDash \alpha$ if and only if $\delta(\sigma_q) \vDash \alpha(\sigma)$. Furthermore $S(\delta(\sigma_q)) = \sigma_q(S(\delta))$ by the observations in the above definition.
Hence $\delta \in SD_q(L)$ is in $D(\alpha, E)$ if and only if $\delta(\sigma_q)$ is in $D(\alpha(\sigma), \sigma_q(E))$. Hence $|D(\alpha, E)| = |D(\alpha(\sigma), \sigma_q(E))|$. And by the lemma above since $S(\beta) = S(\alpha(\sigma))$ we have $|D(\alpha, E)| = |D(\beta, \sigma_q(E))|$. This now yields

$$\begin{aligned} |G(\alpha, \vec{x})| &= \sum_{\substack{E \in ER_q \\ \overrightarrow{\mathcal{S}(E)} = \vec{x}}} |D(\alpha, E)| = \sum_{\substack{E \in ER_q \\ \overrightarrow{\mathcal{S}(E)} = \vec{x}}} |D(\beta, \sigma_q(E))| \\ &= \sum_{\substack{E \in ER_q \\ \overrightarrow{\mathcal{S}(E)} = \vec{x}}} |D(\beta, E)| = |G(\beta, \vec{x})| \ . \end{aligned}$$

Since the spectrum of E and $\sigma_q(E)$ are the same and since summing over all equivalence relation is as good as summing over all equivalence relations permuted with a fixed $\sigma_q \in S^q$. $\square$

Definition 35. Let $\alpha \in SD_p(L)$ have spectrum $\vec{x} \in X_p$. Then define

$$N(\vec{x}, \vec{y}, L) := |\{\beta \in SD(L) \mid \beta \vDash \alpha \text{ and } \overrightarrow{\mathcal{S}(\beta)} = \vec{y}\}| \ . \tag{5.10}$$

Due to the previous result $N(\vec{x}, \vec{y}, L)$ only depends on the spectrum of α and not on α itself, so it is well defined. If $\vec{y} \in X_{<p}$ then $N(\vec{x}, \vec{y}, L) = 0$.
Let furthermore $N(t, L) := N(\emptyset, 1@t, L) = |SD_t^t(L)|$. So $N(\emptyset, 1@t, L)$ counts the number of state descriptions with spectrum $1@t$.
If it's clear (or irrelevant) over which language we are working the explicit mentioning of the language in the definition of N will be omitted.
Let furthermore, provided we don't divide by zero,

$$d(\vec{y}, L, L') := \frac{N(\emptyset, \vec{y}, L)}{N(\emptyset, \vec{y}, L')} \ . \tag{5.11}$$

$N(\emptyset, \vec{y}, L')$ is always greater than zero if $r_0(L') \geq 2$. If the language is purely unary then $N(\emptyset, \vec{y})$ is zero if and only if $|\vec{y}| \geq 2^{m_1} + 1$ since in that case there are not enough atoms to construct a state description with spectrum longer than 2^{m_1}.

Lemma 16. For $\vec{y} \in X^t$ we have, where defined,

$$d(\vec{y}, L, L') = \frac{N(t, L)}{N(t, L')} \ . \tag{5.12}$$

Hence $d(\vec{y}, L, L')$ only depends on t but not on the actual $\vec{y} \in X^t$. We can hence unambiguously write $d(t, L, L')$.

Proof. We use the one to one correspondence between state descriptions and their canonical representations and assume that $N(\emptyset, \vec{y}, L),\ N(\emptyset, \vec{y}, L')$ are not zero.
By Corollary 1 on page 38 we have for any language L that $N(\emptyset, \vec{y}, L) = \binom{p}{\vec{y}} \frac{|SD_t^t(L)|}{\prod_{i=1}^p q_i!}$.
Hence, where defined, we have

$$\begin{aligned} d(\vec{y}, L, L') &= \binom{p}{\vec{y}} \frac{|SD_t^t(L)|}{\prod_{i=1}^p q_i!} \cdot \left[\binom{p}{\vec{y}} \frac{|SD_t^t(L')|}{\prod_{i=1}^p q_i!} \right]^{-1} \\ &= \frac{|SD_t^t(L)|}{|SD_t^t(L')|} = \frac{N(t, L)}{N(t, L')} \ . \end{aligned}$$

□

5.2 Homogeneity and Heterogeneity

Definition 36. Let w satisfy **(SX)**. We say that w is *$\leq t$-heterogeneous* if $w(\vec{x}) = 0$ for all $\vec{x} \in X^{\geq t+1}$.
Furthermore w is called *$t-$heterogeneous* if in addition

$$\lim_{p\to\infty} \sum_{\vec{x}\in X_p^{\leq t-1}} N(\emptyset, \vec{x}) \cdot w(\vec{x}) = \lim_{p\to\infty} \sum_{\alpha\in SD_p^{\leq t-1}} w(\alpha) = 0 \ .$$

In other words w is $t-$heterogeneous if *in the limit* all the probability is massed on the spectra of length exactly t. On a not purely unary language a $\leq t-$heterogeneous probability function can't satisfy **(REG)**.
For a given language there is only one $1-$heterogeneous probability function henceforth denoted by $w^{[1]}$. For which we have $w^{[1]}(\langle p\rangle) = \frac{1}{|SD_1|} = \prod_{d=1}^{r_0} 2^{-m_d}$ and $0 = w^{[1]}(\vec{x})$ for all spectra $\vec{x}$ in $X^{\geq 2}$.

Definition 37. A probability function w satisfying **(SX)** is called *homogeneous* if for all t

$$\lim_{p\to\infty} \sum_{\vec{x}\in X_p^{t}} N(\emptyset, \vec{x}) \cdot w(\vec{x}) = \lim_{p\to\infty} \sum_{\alpha\in SD_p^{t}} w(\alpha) = 0 \ .$$

As we are going to see [Theorem 10 on page 72 and Theorem 11 on page 73] all homogeneous probability functions satisfy **(REG)**.
If L is purely unary then there aren't any homogeneous functions since every probability function has to concentrate all measure on state descriptions with at most 2^{m_1} classes. Since in that case our language does not allow us to tell more constants apart.

5.3 Forth

In this section we investigate close connections between $t-$heterogenous probability functions on purely unary languages and their counterparts on polyadic languages and prove the first half of a representation theorem.

Theorem 3. *Let L' be purely unary with m_1' many relation symbols and L be another language that is not purely unary. Let w be a $t-$heterogeneous probability function on*

L such that $t \leq 2^{m'_1}$. Define for $\alpha \in SD(L')$ with spectrum $\vec{x} \in X$

$$v(\vec{x}) := \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} d(t,L,L') \cdot N(\vec{x},\vec{y},L') \cdot w(\vec{y}) \ . \tag{5.13}$$

Then v extends uniquely to a $t-$heterogeneous probability function on SL'.

Proof. For $t \leq 2^{m'_1}$ $SD_t^t(L')$ is not empty so $d(t,L,L')$ is defined in (5.13) and hence the map v from the set of spectra into the reals is well defined once we know that the limit exists. Now let's prove the existence of the limit first that will make life a lot easier once this is done.

Suppose that $\vec{x} \in X_p^s, \vec{y} \in X_q^s$ with $p \leq q$ and let L_0 be any language in our sense. Due to splitting and joining Lemma 12 on page 41 every new constant in a state description $\beta \in SD_q(L_0)$ with spectrum $\vec{y}$ extending an $\alpha \in SD_p(L_0)$ with spectrum $\vec{x}$ has to join an already existing class.

So for a fixed equivalence relation F with spectrum $\vec{x} \in X_p^s$

$$N(\vec{x},\vec{y},L) = |\{E \in ER_q^s \mid \overrightarrow{\mathcal{S}(E)} = \vec{y}, F = E\restriction_p\}| = N(\vec{x},\vec{y},L') \ . \tag{5.14}$$

Of course this is independent of which equivalence relation F with spectrum $\vec{x}$ we chose.

Next let $p' \leq p < q$ and $\vec{x} \in X_{p'}$. Then

$$\begin{aligned}
&\sum_{\vec{y}\in X_p^t} d(t,L,L') \cdot N(\vec{x},\vec{y},L') \cdot w(\vec{y}) \\
&= \sum_{\vec{y}\in X_p^t} \sum_{\vec{z}\in X_q^t} d(t,L,L') \cdot N(\vec{x},\vec{y},L') \cdot N(\vec{y},\vec{z},L) \cdot w(\vec{z}) && (5.15)\\
&= \sum_{\vec{y}\in X_p^t} \sum_{\vec{z}\in X_q^t} d(t,L,L') \cdot N(\vec{x},\vec{y},L') \cdot N(\vec{y},\vec{z},L') \cdot w(\vec{z}) && (5.16)\\
&\leq \sum_{\vec{z}\in X_q^t} d(t,L,L') \cdot N(\vec{x},\vec{z},L') \cdot w(\vec{z}) \ . && (5.17)
\end{aligned}$$

Equation (5.15) is due to the fact that w is a $t-$heterogeneous probability function.
(5.16) follows from (5.14).
The inequality follows from the following observation. Let $k \leq n \leq m \in \mathbb{N}$ then for

$\vec{x} \in X_k$ and $\vec{z} \in X_m$ and any language L_0

$$N(\vec{x}, \vec{z}, L_0) = \sum_{\vec{y} \in X_n} N(\vec{x}, \vec{y}, L_0) \cdot N(\vec{y}, \vec{z}, L_0) \tag{5.18}$$

$$\geq \sum_{\vec{y} \in X_n^t} N(\vec{x}, \vec{y}, L_0) \cdot N(\vec{y}, \vec{z}, L_0) \ . \tag{5.19}$$

So the sequence defining $v(\vec{x})$ increases as p gets ever bigger. For fixed p we have

$$\begin{aligned} 0 &\leq \sum_{\vec{y} \in X_p^t} d(t, L, L') \cdot N(\vec{x}, \vec{y}, L') \cdot w(\vec{y}) \\ &\leq \sum_{\vec{y} \in X_p^t} d(t, L, L') \cdot N(\emptyset, \vec{y}, L') \cdot w(\vec{y}) \\ &= \sum_{\vec{y} \in X_p^t} N(\emptyset, \vec{y}, L) \cdot w(\vec{y}) \leq 1 \ . \end{aligned}$$

So the sequence defining $v(\vec{x})$ is not only non-decreasing but also bounded from above by one. Hence the limit exists.
For any probability function w_0 on any language L_0 satisfying **(SX)** we have for $\vec{x} \in X_p$ and $q \geq p$

$$w_0(\vec{x}) = \sum_{\vec{y} \in X_q} N(\vec{x}, \vec{y}, L_0) \cdot w_0(\vec{y}) \ .$$

Since w is $t-$heterogeneous this implies

$$w(\vec{x}) = \lim_{q \to \infty} \sum_{\vec{y} \in X_q^t} N(\vec{x}, \vec{y}, L) \cdot w(\vec{y}) \ .$$

For all $\vec{x} \in X$ and all languages L_0

$$N(\emptyset, \vec{x}, L_0) = N(\emptyset, \langle 1 \rangle, L_0) \cdot \ N(\langle 1 \rangle, \vec{x}, L_0) \ . \tag{5.20}$$

That is a special case of equation (5.18) as X_1 has only one element. Since w is $t-$heterogeneous we now find

$$N(\emptyset, \langle 1 \rangle, L') \cdot v(\langle 1 \rangle) = \lim_{p \to \infty} \sum_{\vec{x} \in X_p^t} \frac{N(\emptyset, \vec{x}, L)}{N(\emptyset, \vec{x}, L')} N(\emptyset, \langle 1 \rangle, L') \cdot N(\langle 1 \rangle, \vec{x}, L') \cdot w(\vec{x})$$

$$= \lim_{p \to \infty} \sum_{\vec{x} \in X_p^t} N(\emptyset, \vec{x}, L) \cdot w(\vec{x}) = 1 \ .$$

Furthermore we find for all q and all $\vec{x} \in X_q^{\leq t}$ by (5.18)

$$\begin{aligned} v(\vec{x}) &= \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} d(t,L,L')\cdot N(\vec{x},\vec{y},L')\cdot w(\vec{y}) \\ &= \lim_{p\to\infty} \sum_{\vec{z}\in X_{q+1}} N(\vec{x},\vec{z},L') \sum_{\vec{y}\in X_p^t} d(t,L,L')\cdot N(\vec{z},\vec{y},L')\cdot w(\vec{y}) \\ &= \sum_{\vec{z}\in X_{q+1}} N(\vec{x},\vec{z},L')\cdot v(\vec{z}) \ \ . \end{aligned}$$

Now we can apply Lemma 4 on page 18 to see that v extends to a probability function. It's clear that v satisfies **(SX)**. The $t-$heterogeneity is now the only claim that needs checking.
Now fix q then

$$\sum_{\vec{x}\in X_q^t} N(\emptyset,\vec{x},L')\cdot v(\vec{x})$$

$$= \sum_{\vec{x}\in X_q^t} N(\emptyset,\vec{x},L')\cdot \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} d(t,L,L')\cdot N(\vec{x},\vec{y},L')\cdot w(\vec{y}) \tag{5.21}$$

$$= \sum_{\vec{x}\in X_q^t} \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} N(\emptyset,\vec{x},L')\cdot \frac{N(\emptyset,\vec{x},L)}{N(\emptyset,\vec{x},L')}\cdot N(\vec{x},\vec{y},L')\cdot w(\vec{y}) \tag{5.22}$$

$$= \sum_{\vec{x}\in X_q^t} N(\emptyset,\vec{x},L)\cdot \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} N(\vec{x},\vec{y},L)\cdot w(\vec{y}) \tag{5.23}$$

$$= \sum_{\vec{x}\in X_q^t} N(\emptyset,\vec{x},L)\cdot w(\vec{x}) \ \ . \tag{5.24}$$

Equation (5.23) is due to the fact that $\vec{x}$ and $\vec{y}$ have the same length. Hence can we make use of (5.14). The last equality holds because w is $t-$heterogeneous.
The last sum tends to one as q tends to infinity since w is $t-$heterogeneous. Hence v is $t-$heterogeneous. □

5.4 and back again

Theorem 4. *Let v be a $t-$heterogeneous probability function on a purely unary language L'. Suppose L is any not purely unary language. If we define w on the spectra*

$\vec{x}$ of state descriptions of L by

$$w(\vec{x}) := \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} d(t, L', L) \cdot N(\vec{x}, \vec{y}, L) \cdot v(\vec{y}) \ , \tag{5.25}$$

then w extends uniquely to a $t-$heterogeneous probability function on SL. [4]

Proof. Here $d(t, L', L)$ is always defined since $N(\emptyset, \vec{x}, L) > 0$ for all $x \in X$.
We now swap L and L' and w with v in the proof above and we can use the above proof here. □

Corollary 2. The operations in the two theorems above are inverse to each other.

Proof. Since w and is $t-$heterogeneous it holds for all $\vec{x} \in X^{\leq t}$

$$\begin{aligned} w(\vec{x}) &= \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} N(\vec{x}, \vec{y}, L) w(\vec{y}) \\ &= \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} N(\vec{x}, \vec{y}, L) \cdot \lim_{q\to\infty} \sum_{\vec{z}\in X_q^t} N(\vec{y}, \vec{z}, L') \cdot w(\vec{z}) \\ &= \lim_{p\to\infty} \sum_{\vec{y}\in X_p^t} N(\vec{x}, \vec{y}, L) \cdot d(t, L', L) \cdot \lim_{q\to\infty} \sum_{\vec{z}\in X_q^t} d(t, L, L') \cdot N(\vec{y}, \vec{z}, L') \cdot w(\vec{z}). \end{aligned}$$

The other implication of the corollary follows from the fact that the equations above also hold if we swap v with w and L with L'. □

5.5 The $\eta-$Representation

We will now see that we can uniquely decompose every probability function w satisfying **(SX)** in heterogeneous parts and a homogeneous part.

In the proof below we start by subtracting *the* $1-$heterogeneous part from w. Modulo scale factors w minus the $1-$heterogeneous part is a probability function which we will denote by $w^{(1)}$.
As it turns out this procedure can be iterated. We will subtract *the* $t-$heterogeneous part $w^{[t]}$ from $w^{(t-1)}$ yielding $w^{(t)}$ again up to some factors that are given explicitly in the proof.

[4] Technically $v(\vec{y})$ is not defined for $\vec{y}$ with too long a spectrum as there is no state description over L' with this spectrum. We here simply let for such $\vec{y}$ $v(\vec{y}) := 0$.

After a countable number of steps we are either left empty handed or with a homogeneous probability function $w^{[0]}$ (modulo some factors).
The dear reader should not lose heart in face of such a long proof. Most of space is devoted to calculations checking that the defined functions are indeed probability functions and proving the here outlined propositions.
In this section we will frequently use the following fact for $\vec{x} \in X_p$ and $\vec{y} \in X_q$ with $p \leq r \leq q$

$$N(\vec{x}, \vec{y}) = \sum_{\vec{z} \in X_r} N(\vec{x}, \vec{z}) N(\vec{z}, \vec{y}) \ .$$

We have encountered the equality before in equation (5.18) on page 49.

Theorem 5. *Let L be not purely unary and assume that w satisfy **(SX)**. Then there are $t-$heterogeneous probability functions $w^{[t]}$, constants η_t for $t \geq 1$, a constant η_0 and a homogeneous probability function $w^{[0]}$ such that*

$$w = \sum_{i=0}^{\infty} \eta_i w^{[i]} \quad \textit{and} \quad \sum_{i=0}^{\infty} \eta_i = 1 \ . \tag{5.26}$$

The η_i are unique and so are the $w^{[i]}$ whenever $\eta_i \neq 0$ furthermore $\eta_i \in [0, 1]$ for all $i \geq 0$.
*Conversely every such infinite convex combination of $i-$heterogeneous and possibly a homogeneous probability function satisfies **(SX)**.*

Proof. Let

$$\begin{aligned} \gamma_1 : &= \lim_{p \to \infty} \sum_{\vec{x} \in X_p^1} N(\emptyset, \vec{x}) w(\vec{x}) = \lim_{p \to \infty} N(\emptyset, \langle p \rangle) w(\langle p \rangle) \\ &= \lim_{p \to \infty} N(\emptyset, \langle 1 \rangle) w(\langle p \rangle) = |SD_1| \cdot \lim_{p \to \infty} w(\langle p \rangle) \ . \end{aligned} \tag{5.27}$$

This limit exists since it's bounded from below by zero and the sequence is decreasing, not necessarily strictly, as p gets ever bigger.
If $\gamma_1 \neq 0$ put

$$w^{[1]}(\vec{x}) := \lim_{k \to \infty} \sum_{\vec{y} \in X_k^1} \gamma_1^{-1} N(\vec{x}, \vec{y}) w(\vec{y}) = \lim_{k \to \infty} \gamma_1^{-1} N(\vec{x}, \langle k \rangle) w(\langle k \rangle) \ .$$ [5]

[5] At this point there's a small clash of notation as we have $w^{[1]}$ already defined as the unique $1-$heterogeneous function. As we are going to see shortly the here defined $w^{[1]}$ is the same as the $1-$heterogeneous function.

Observe that $w^{[1]}$ gives value zero to spectra $\vec{x}$ that don't have length one since in that case $N(\vec{x},\vec{y}) = 0$ for $\vec{y} \in X^1$ and $\vec{x} \in X^{\geq 2}$.
Furthermore for every $q \in \mathbb{N}$

$$w^{[1]}(\langle q\rangle) = \frac{\lim_{k\to\infty} N(\langle q\rangle,\langle k\rangle)w(\langle k\rangle)}{|SD_1|\cdot\lim_{p\to\infty} w(\langle p\rangle)} = |SD_1|^{-1} = w^{[1]}(\langle 1\rangle) \ .$$

So, as advertised, $w^{[1]}$ is indeed what it should be.
If $\gamma_1 = 1$ then $w = w^{[1]}$ and we are done. Because then w concentrates all probability on spectra with length 1. So assume from now on $\gamma_1 < 1$.
Suppose that $\gamma_1 < 1$ then define $w^{(1)} := \begin{cases} (1-\gamma_1)^{-1}(w-\gamma_1 w^{[1]}) & \text{if } \gamma_1 \neq 0 \\ w & \text{if } \gamma_1 = 0. \end{cases}$
For $\gamma_1 < 1$ we have for all $\vec{y} \in X^1$ that $w(\vec{y}) \geq \lim_{p\to\infty} w(\langle p\rangle) = \gamma_1 \cdot w^{[1]}(\vec{y})$ and for all $\vec{z} \in X^{\geq 2}$ that $w(\vec{z}) \geq 0 = \gamma_1 w^{[1]}(\vec{z})$. Hence for all $\alpha \in SD$ we have $w(\alpha) - \gamma_1 w^{[1]}(\alpha) \geq 0$.
Furthermore for all $\vec{x} \in X_p$ we have $w^{(1)}(\vec{x}) = \sum_{\vec{y}\in X_{p+1}} N(\vec{x},\vec{y})w^{(1)}(\vec{y})$ and we have $N(\emptyset,\langle 1\rangle)w^{(1)}(\langle 1\rangle) = 1$. Hence $w^{(1)}$ is a probability function by our standard lemma (see page 18 Lemma 4) and satisfies **(SX)**.
Furthermore we have

$$w = \gamma_1 w^{[1]} + (1-\gamma_1)w^{(1)} \ . \tag{5.28}$$

Next we want to see that $w^{(1)}$ does not contain a $1-$heterogeneous part.
Assuming $\gamma_1 > 0$ we have that since $w^{[1]}$ is $1-$heterogeneous

$$\begin{aligned}
\lim_{p\to\infty}\sum_{\vec{x}\in X^1_p} N(\emptyset,\vec{x})w^{(1)}(\vec{x}) &= \lim_{p\to\infty} N(\emptyset,\langle p\rangle)\frac{w(\langle p\rangle)-\gamma_1 w^{[1]}(\langle p\rangle)}{1-\gamma_1} \\
&= \frac{\gamma_1 - \lim_{p\to\infty} N(\emptyset,\langle p\rangle)\gamma_1 w^{[1]}(\langle p\rangle)}{1-\gamma_1} \\
&= \frac{\gamma_1 - \gamma_1\cdot\lim_{p\to\infty} N(\emptyset,\langle 1\rangle)w^{[1]}(\langle 1\rangle)}{1-\gamma_1} \\
&= \frac{\gamma_1 - \gamma_1\cdot N(\emptyset,\langle 1\rangle)w^{[1]}(\langle 1\rangle)}{1-\gamma_1} \\
&= \frac{\gamma_1-\gamma_1\cdot 1}{1-\gamma_1} = 0 \ .
\end{aligned}$$

If $\gamma_1 = 0$ that means that w didn't contain *any* $1-$heterogeneous part. In this case

$w = w^{(1)}$ and so

$$\lim_{p\to\infty} \sum_{\vec{x}\in X_p^1} N(\emptyset,\vec{x})w^{(1)}(\vec{x}) = \lim_{p\to\infty} \sum_{\vec{x}\in X_p^1} N(\emptyset,\vec{x})w(\vec{x}) = \gamma_1 = 0 \ .$$

Now for the **inductive step** assume we have defined probability functions $w^{(j)}$ for $1 \le j \le l-1$ satisfying **(SX)** and all γ_j defined so far are less than one. Also assume that $w^{(l-1)}$ puts in the limit no measure on spectra of length $l-1$ or less and that $w^{[j]}$ is $j-$heterogeneous for $1 \le j \le l-1$.
Then let

$$\gamma_l := \lim_{p\to\infty} \sum_{\vec{x}\in X_p^l} N(\emptyset,\vec{x})w^{(l-1)}(\vec{x}) \text{ [6]}$$

and assuming $0 < \gamma_l$ we put

$$w^{[l]}(\vec{x}) := \lim_{p\to\infty} \sum_{\vec{y}\in X_p^l} \gamma_l^{-1} N(\vec{x},\vec{y})w^{(l-1)}(\vec{y}) \ . \tag{5.29}$$

We will have to check that this defines an $l-$heterogeneous function $w^{[l]}$ for $\gamma_l > 0$.
If $\gamma_l = 1$ then $w^{(l-1)}$ is $l-$heterogeneous and our inductive process terminates by setting $\eta_{l+n} = \eta_0 = 0$ for all n greater or equal than one.
If $\gamma_l < 1$ we let

$$w^{(l)} := \begin{cases} (1-\gamma_l)^{-1}(w^{(l-1)} - \gamma_l w^{[l]}) & \text{if } \gamma_l \neq 0 \\ w^{(l-1)} & \text{if } \gamma_l = 0. \end{cases} \tag{5.30}$$

Using this inductive definition for $w^{(l-1)}$ and that the $w^{[j]}$ are $j-$heterogeneous we note for later use

$$\gamma_l = \lim_{p\to\infty} \sum_{\vec{x}\in X_p^l} N(\emptyset,\vec{x})w^{(l-1)}(\vec{x}) = \lim_{p\to\infty} \sum_{\vec{x}\in X_p^l} N(\emptyset,\vec{x})w(\vec{x}) \prod_{i=1}^{l-1} \frac{1}{1-\gamma_i} \ .\text{[7]} \tag{5.31}$$

Let's now check that $w^{[l]}$ is a $l-$heterogeneous probability function, therefore compute

[6] The limit exists for the same reason the limit in the definition of γ_1 exists and since we assumed that $\lim_{p\to\infty} \sum_{\vec{x}\in X_p^{\le l-1}} N(\emptyset,\vec{x})w^{(l-1)}(\vec{x}) = 0$.

[7] By the induction hypothesis none of the γ_i equals one, hence we never divide by zero.

at first

$$N(\emptyset,\langle 1\rangle)\cdot\ w^{[l]}(\langle 1\rangle) = N(\emptyset,\langle 1\rangle)\lim_{p\to\infty}\sum_{\vec{x}\in X_p^l}\gamma_l^{-1}N(\langle 1\rangle,\vec{x})\ w^{(l-1)}(\vec{x})$$
$$= \gamma_l^{-1}\lim_{p\to\infty}\sum_{\vec{x}\in X_p^l}N(\emptyset,\vec{x})\ w^{(l-1)}(\vec{x}) = 1$$

by the definition of γ_l. For all $\vec{x}\in X_p^{\leq l}$ we find by twice using the definition of $w^{[l]}$

$$\begin{aligned} w^{[l]}(\vec{x}) &= \lim_{q\to\infty}\sum_{\vec{z}\in X_q^l}\gamma_l^{-1}N(\vec{x},\vec{z})w^{(l-1)}(\vec{z}) \\ &= \lim_{q\to\infty}\sum_{\vec{z}\in X_q^l}\sum_{\vec{y}\in X_{p+1}}\gamma_l^{-1}N(\vec{x},\vec{y})N(\vec{y},\vec{z})w^{(l-1)}(\vec{z}) \\ &= \sum_{\vec{y}\in X_{p+1}}N(\vec{x},\vec{y})\lim_{q\to\infty}\sum_{\vec{z}\in X_q^l}\gamma_l^{-1}N(\vec{y},\vec{z})w^{(l-1)}(\vec{z}) \\ &= \sum_{\vec{y}\in X_{p+1}}N(\vec{x},\vec{y})w^{[l]}(\vec{y}) \end{aligned}$$

and finally

$$\lim_{p\to\infty}\sum_{\vec{x}\in X_p^l}N(\emptyset,\vec{x})w^{[l]}(\vec{x}) = \lim_{p\to\infty}\sum_{\vec{x}\in X_p^l}N(\emptyset,\vec{x})\lim_{q\to\infty}\sum_{\vec{y}\in X_q^l}\gamma_l^{-1}N(\vec{x},\vec{y})w^{(l-1)}(\vec{y})$$
$$= \lim_{q\to\infty}\sum_{\vec{y}\in X_q^l}\gamma_l^{-1}N(\emptyset,\vec{y})w^{(l-1)}(\vec{y}) = \frac{\gamma_l}{\gamma_l} = 1\ .$$

So $w^{[l]}$ defines a probability function by Lemma 4 on page 18. By the last equation it is also $l-$heterogeneous.
By the definition of $w^{[l]}$ and $w^{(l)}$ we have $w^{(l-1)}(\vec{x}) - \gamma_l\cdot w^{[l]}(\vec{x}) \geq 0$ for all $\vec{x}\in X$. Furthermore for $\vec{x}\in X_p$ we have $w^{(l)}(\vec{x}) = \sum_{\vec{y}\in X_{p+1}}N(\vec{x},\vec{y})w^{(l)}$ since this holds for $w^{(l-1)}$ and $w^{[l]}$. Since $w^{[l]}$ and $w^{(l-1)}$ are probability functions we have for $1>\gamma_l>0$ that

$$\sum_{\alpha\in SD_1}w^{(l)}(\alpha) = \frac{\sum_{\alpha\in SD_1}w^{(l-1)}(\alpha)-\gamma_l\cdot w^{[l]}(\alpha)}{1-\gamma_l} = \frac{1-\gamma_l\cdot\sum_{\alpha\in SD_1}w^{[l]}(\alpha)}{1-\gamma_l} = 1\ .$$

So $w^{(l)}$ is a probability function for $0\leq\gamma_l<1$.
By our induction hypothesis $w^{(l-1)}$ has in the limit no measure on spectra of length

strictly less than l hence

$$\lim_{p\to\infty}\sum_{\vec{x}\in X_p^{\leq l}} N(\emptyset,\vec{x})w^{(l)}(\vec{x}) = \lim_{p\to\infty}\sum_{\vec{x}\in X_p^{\leq l}} N(\emptyset,\vec{x})\frac{w^{(l-1)}(\vec{x})-\gamma_l w^{[l]}(\vec{x})}{1-\gamma_l} \tag{5.32}$$

$$= \lim_{p\to\infty}\sum_{\vec{x}\in X_p^{l}} N(\emptyset,\vec{x})\frac{w^{(l-1)}(\vec{x})-\gamma_l w^{[l]}(\vec{x})}{1-\gamma_l} \tag{5.33}$$

$$= \frac{\gamma_l - \gamma_l \lim_{p\to\infty}\sum_{\vec{x}\in X_p^l} N(\emptyset,\vec{x})w^{[l]}(\vec{x})}{1-\gamma_l} \tag{5.34}$$

$$= \frac{\gamma_l-\gamma_l}{1-\gamma_l} \quad = \quad 0 \ . \tag{5.35}$$

What proves that $w^{(l)}$ puts in the limit no measure on all spectra in $X_p^{\leq l}$.
The η_l are computed from (5.28) and (5.30) which gives for $1 \leq j < \infty$

$$\eta_j := \gamma_j \cdot \prod_{i=1}^{j-1}(1-\gamma_i) \tag{5.36}$$

since by our definitions

$$w = \eta_1 w^{[1]} + \cdots + \eta_l w^{[l]} + (1 - \sum_{i=1}^{l}\eta_i)\cdot w^{(l)} \ . \tag{5.37}$$

Clearly $0 \leq \sum_{j=1}^{\infty}\eta_j \leq 1$ since $\sum_{l=1}^{\infty}\lim_{p\to\infty}\sum_{\alpha\in SD_p^l} w(\alpha) \leq 1$.
Then we can define $\eta_0 := 1 - \sum_{i=1}^{\infty}\eta_i$. Supposing that η_0 is greater than zero we then have to see that $w^{[0]}$ is a probability function where $w^{[0]}$ is defined by

$$\eta_0 \cdot w^{[0]} := w - \sum_{i=1}^{\infty}\eta_i w^{[i]} \ .$$

By (5.37) this is never less than zero. For all p and all $\vec{x}\in X_p$ we have that

$$\sum_{i=1}^{\infty}\eta_i w^{[i]}(\vec{x}) = \sum_{\vec{y}\in X_{p+1}} N(\vec{x},\vec{y})\sum_{i=1}^{\infty}\eta_i w^{[i]}(\vec{y})$$

since it holds for all $w^{[i]}$.

Furthermore since this equation also holds for the probability function w we have

$$w^{[0]}(\vec{x}) = \sum_{\vec{y}\in X_{p+1}} N(\vec{x},\vec{y})w^{[0]}(\vec{y}) \ .$$

The idea is to use Lemma 4 on page 18 to show that $w^{[0]}$ is indeed a probability function.

Since

$$1 = \sum_{\alpha\in SD_1} w(\alpha) = \sum_{\alpha\in SD_1}\sum_{i=0}^{\infty} \eta_i \cdot w^{[i]}(\alpha) = \Big(\sum_{i=1}^{\infty}\eta_i\Big) + \Big(\sum_{\alpha\in SD_1} \eta_0 \cdot w^{[0]}(\alpha)\Big)$$

we have

$$1 - \sum_{i=1}^{\infty}\eta_i = \eta_0 \cdot \sum_{\alpha\in SD_1} w^{[0]}(\alpha) \ .$$

Substituting the definition of η_0 gives $1 = \sum_{\alpha\in SD_1} w^{[0]}(\alpha)$. So $w^{[0]}$ is indeed a probability function by our standard lemma.

$w^{[0]}$ satisfies **(SX)** since all the $w^{[l]}$ and w do and furthermore fulfilling **(SX)** is a closed condition, so the limit in $\eta_0 w^{[0]} = w - \lim_{l\to\infty}\sum_{i=1}^{l}\eta_l w^{[l]}$ poses no problem.

The only thing left to prove is the homogeneity of $w^{[0]}$ for $\eta_0 > 0$.

Assume that u is minimal such that $\lim_{p\to\infty}\sum_{\vec{p}\in X_p^u} N(\emptyset,\vec{p})\cdot w^{[0]}(\vec{p}) \neq 0$. If there is no such u then $w^{[0]}$ is homogeneous. We'll assume that such an u exists and derive a contradiction.

Now let $\epsilon > 0$ and pick a $k > u$ such that $\sum_{i>k}^{\infty}\eta_i < \epsilon$.

Since the $w^{[r]}$ are r-heterogeneous we have that for all $\epsilon' > 0$ there exists a p_0 such that for all $p' > p_0$ and $u < r \leq k$ we have $\sum_{\vec{x}\in X_{p'}^u} N(\emptyset,\vec{x})w^{[r]}(\vec{x}) < \epsilon'$.

Then

$$\sum_{\vec{x}\in X_{p'}^u}\sum_{i=u+1}^{\infty} N(\emptyset,\vec{x})\eta_i w^{[i]}(\vec{x}) = \sum_{i=u+1}^{\infty}\eta_i \sum_{\vec{x}\in X_{p'}^u} N(\emptyset,\vec{x})w^{[i]}(\vec{x}) < \epsilon'(k-u) + \sum_{i>k}\eta_i \ .$$

Now if $\epsilon' = \epsilon/(k-u)$ then this sum is smaller than $2\cdot\epsilon$. So

$$\lim_{p\to\infty}\sum_{\vec{x}\in X_p^u} N(\emptyset,\vec{x})\sum_{i=u+1}^{\infty}\eta_i w^{[i]}(\vec{x}) = 0 \ . \tag{5.38}$$

Prepared with these observations we'll see that $w^{[0]}$ is homogeneous

$$\begin{aligned}
&\lim_{p\to\infty} \sum_{\vec{x}\in X_p^u} N(\emptyset,\vec{x}) \cdot \eta_0 w^{[0]}(\vec{x}) \\
&= \lim_{p\to\infty} \sum_{\vec{x}\in X_p^u} N(\emptyset,\vec{x}) \cdot \Big(w(\vec{x}) - \sum_{i=1}^{\infty} \eta_i w^{[i]}(\vec{x}) \Big) \\
&= \lim_{p\to\infty} \sum_{\vec{x}\in X_p^u} N(\emptyset,\vec{x}) \cdot \Big(w(\vec{x}) - \sum_{i=1}^{u} \eta_i w^{[i]} \vec{x}) - \sum_{i=u+1}^{\infty} \eta_i w^{[i]}(\vec{p}) \Big) \\
&= \lim_{p\to\infty} \sum_{\vec{x}\in X_p^u} N(\emptyset,\vec{x}) \cdot \Big((1 - \sum_{i=1}^{u} \eta_i) w^{(u)}(\vec{x}) - \sum_{i=u+1}^{\infty} \eta_i w^{[i]}(\vec{x}) \Big) \quad \text{using (5.37)} \\
&= \lim_{p\to\infty} \sum_{\vec{x}\in X_p^u} N(\emptyset,\vec{x}) \cdot \Big((1 - \sum_{i=1}^{u} \eta_i) w^{(u)}(\vec{x}) \Big) \quad \text{because of (5.38)} \\
&= 0 \qquad \text{by the equations beginning with (5.32)}
\end{aligned}$$

which is a contradiction to our assumption. So $w^{[0]}$ is homogeneous if $\sum_{i=1}^{\infty} \eta_i < 1$, that is $\eta_0 \neq 0$.
That such a sum of probability functions satisfies **(SX)** follows from the fact that satisfying **(SX)** is a closed condition and all the $w^{[i]}$ are assumed to satisfy **(SX)** and the fact that the η_i are all in $[0, 1]$ and sum to one. □

We want to mention that is well known that Theorem 5 also holds for purely unary languages L. If L is purely unary then $\sum_{i=1}^{2^{m_1}} \eta_i$ has to equal one since we can at most tell 2^{m_1} many constants over L apart. All other η_i are zero.

Chapter 6

Sublanguages and (SX)

Theorem 6. *Let L' be a sublanguage of L and let w be a probability function on SL that satisfies **(SX)**. Then the restriction of w to SL' w' satisfies **(SX)**.*

Proof. Consider $\alpha, \beta \in SD_p(L')$ with the same spectrum.[1] We will see that $w'(\alpha) = w(\alpha) = w(\beta) = w'(\beta)$ holds. The first and the last equality are trivially satisfied since w agrees with w' on L'.
Since w satisfies **(SX)** the probability of sentences of L is invariant under renaming constants. Since L' is a sublanguage of L the probability of sentences of L' is invariant under renaming constants. So we can assume that $S(\alpha) = S(\beta)$.
Note that then for all $\gamma \in SD_p(L \setminus L')$ $\overrightarrow{\mathcal{S}(\alpha \wedge \gamma)} = \overrightarrow{\mathcal{S}(\beta \wedge \gamma)}$ and hence

$$w'(\alpha) = w(\alpha) = \sum_{\gamma \in SD_p(L \setminus L')} w(\alpha \wedge \gamma) = \sum_{\gamma \in SD_p(L \setminus L')} w(\beta \wedge \gamma) = w(\beta) = w'(\beta) \ .$$

□

Corollary 3. If w is $\leq t-$heterogeneous on L then so are all restrictions to sub languages.

Proof. Assume one restriction is not. Then there is sublanguage $L' \subset L$ and $\alpha \in SD_p^{>t}(L')$ such that $w(\alpha) > 0$.
As always $w(\alpha) = \sum_{\beta \in SD_p(L \setminus L')} w(\beta \wedge \alpha.)$ The $\beta \wedge \alpha$ all have a spectrum that is not shorter than that of α. Hence the $\beta \wedge \alpha$ are in $SD^{>t}(L)$. Since we assumed that w is $\leq t-$heterogeneous we have for all $\beta \in SD_p(L \setminus L')$ $w(\beta \wedge \alpha) = 0$. That implies $w(\alpha) = 0$, giving the required contradiction. □

[1]Since α and β are state descriptions over L' $S(\alpha)$ and $S(\beta)$ are well defined.

Chapter 7

De Finetti-Style Representations

7.1 The Heterogeneous Case

The main theorem in this section was first mentioned in [25] for binary languages and later proved in [18] for general polyadic languages. We here fill out some of the missing details and give in fact a different proof from the one given in [18]. For a representation result of probability functions satisfying **(CX)** see [10]. For representation results for *symmetric* probability functions see [14].
Throughout we will assume that L is not purely unary and $t \geq 2$. The case $t = 1$ is trivial. We already know that there is only one $1-$heterogeneous probability function.

First we give some intuition as to how the whole construction works.
Suppose we are given an urn containing t colored balls with pairwise different colors $c_1, \ldots, c_t$. Every color c_i comes with a fixed probability p_i of being drawn such that $\sum_{i=1}^{t} p_i = 1$. Furthermore we assume that for $i < k$ we have $p_i \geq p_k$ and $p_t > 0$. After drawing a ball we put the ball back in the urn; that is we draw *with* replacement. We say that the urn is given by $\vec{p} := \langle p_1, \ldots, p_t \rangle$.
Let B_p the set of all draws of p balls. Define an equivalence relation $S(c)$ for $c \in B_p$ on p by letting $S(c)ik$ if and only if the $i-$th and the $k-$th ball in c have the same color. Note that $|S(c)|$ is the number of different colors in c.
For $c \in B_p$ let $prob(c)$ be the probability of drawing c.
Given such an urn we will inductively construct a $t-$heterogeneous probability function $v^{\vec{p}}$.

First step: Pick one ball from the urn, all choices of balls from the urn are with respect to $\vec{p}$, and choose with the uniform distribution an $\alpha^1 \in SD_1$.

Inductive step from $p-1$ **to** p : Assume we have so far constructed an $\alpha^{p-1} \in SD_{p-1}$ according to a draw in B_{p-1}. Pick another ball from the urn and let c denote the draw of the first p balls. That is we have our first $p-1$ balls and then we draw the $p-$th ball. If the $p-$th ball has the same color as the $i-th$ ball then uniquely extend α^{p-1} to α^p such that $S(\alpha^p)pi$ holds. There is only one such α^p since a_p has to look exactly the same as a_i. If λ is an atomic formula α^p has to decide that is not decided by α^{p-1} we put

$$\alpha^p \models \lambda \text{ if and only if } \alpha^{p-1} \models \lambda(a_p \mapsto a_i) \ .$$

If the $p-$th ball has a color not among the previously chosen colors then α^p extending α^{p-1} has to satisfy that if the $i-$th and the $j-$th ball have the same color then $S(\alpha^p)ij$. Choose such an α^p with probability

$$\frac{N(\overrightarrow{\mathcal{S}(\alpha^p/S(c))}, 1@t)}{N(\overrightarrow{\mathcal{S}(\alpha^{p-1}/S(c))}, 1@t)} \ .^1 \tag{7.1}$$

The probability that for a given draw $c \in B_p$ we will choose a state description $\alpha \in SD_p$ will be denoted by $Cont(\alpha|c)$. We think of it as the contribution of c to the probability of α.

Note that if $S(c)$ is not a refinement of $S(\alpha)$ then $Cont(\alpha|c) = 0$, since two balls of the same color in c lead to equivalent constants in α.

We now let for $\alpha \in SD_p$

$$v_L^{\vec{p}}(\alpha) := \sum_{c \in B_p} prob(c)Cont(\alpha|c) = \sum_{\substack{c \in B_p \\ S(\alpha) \leq S(c)}} prob(c)\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))}, 1@t, L)}{N(\emptyset, 1@t, L)} \ . \tag{7.2}$$

We will drop the subscript L whenever there can't be any confusion.

There's another way of defining $v^{\vec{p}}(\alpha)$. The alternative is

$$v^{\vec{p}}(\alpha) := \sum_{\substack{c \in B_p \\ S(\alpha) \leq S(c)}} \sum_{\substack{\beta \in SD_{|S(c)|} \\ \beta \cdot S(c) = \alpha}} prob(c) \cdot \frac{N(\overrightarrow{\mathcal{S}(\beta)}, 1@t)}{N(\emptyset, 1@t)} \ .$$

[1] We introduced α/E on in definition 29 on page 33. Since $\alpha^p/S(c)$ is a state description we have defined its spectrum.

We want to mention that for a fixed $c \in B_p$ with $S(\alpha) \leq S(c)$ there is only one such $\beta \in SD_{|S(c)|}$. This gives the same probability function as the construction above but we think that the latter definition is rather hard to grasp upon first reading. It does however have advantages in computations.

Theorem 7. *The $v^{\vec{p}}$ are probability functions.*

Proof.

$$\begin{aligned}
\sum_{\alpha \in SD_1} v^{\vec{p}}(\alpha) &= \sum_{\alpha \in SD_1} \sum_{c \in B_1} prob(c) \frac{N(\langle 1 \rangle, 1@t)}{N(\emptyset, 1@t)} \\
&= \sum_{\alpha \in SD_1} \sum_{c \in B_1} prob(c) \frac{N(\langle 1 \rangle, 1@t)}{N(\emptyset, \langle 1 \rangle) N(\langle 1 \rangle, 1@t)} \\
&= \sum_{\alpha \in SD_1} \sum_{c \in B_1} \frac{prob(c)}{|SD_1|} = \sum_{\alpha \in SD_1} \frac{1}{|SD_1|} \cdot \sum_{i=1}^{t} p_i = \sum_{\alpha \in SD_1} \frac{1}{|SD_1|} \\
&= 1 \ .
\end{aligned}$$

We will use the following observation to verify that $\sum_{\beta \in SD_{p+1} \beta \vDash \alpha} v^{\vec{p}}(\beta) = v^{\vec{p}}(\alpha)$ for $\alpha \in SD_p$.
Let $\beta \in SD_{p+1}^{\geq |S(\alpha)|}$ such that $\beta \vDash \alpha$, that is $\beta \restriction_p = \alpha$. Let furthermore $E \in ER_{p+1}$ with $S(\beta) \leq S(E)$ and $p+1 \notin V(E)$, then $p+1$ is not a singleton over $S(\beta)$ and so a_{p+1} is equivalent to some a_i over β for $1 \leq i \leq p$. We hence have

$$\beta / E = \beta \restriction_p / (E \restriction_p) = \alpha / (E \restriction_p) \ . \qquad (*)$$

Since $p+1$ is not new over $S(\beta)$ there is only one $\beta \in SD_{p+1}$ such that $\beta \vDash \alpha$ and $S(\beta) \leq E$.
For $\alpha \in SD_p^l$ with $1 \leq l \leq t$ we now find

$$\begin{aligned}
\sum_{\substack{\beta \in SD_{p+1} \\ \beta \vDash \alpha}} v^{\vec{p}}(\beta) &= \sum_{\substack{\beta \in SD_{p+1} \\ \beta \vDash \alpha}} \sum_{\substack{f \in B_{p+1} \\ S(\alpha) \leq S(f) \restriction_p}} prob(f) \cdot Cont(\beta | f) \\
&= \sum_{\substack{\beta \in SD_{p+1} \\ \beta \vDash \alpha}} \sum_{\substack{f \in B_{p+1} \\ S(\alpha) \leq S(f) \restriction_p \\ p+1 \notin V(S(f))}} prob(f) \cdot Cont(\beta | f) \\
&\quad + \sum_{\substack{\beta \in SD_{p+1} \\ \beta \vDash \alpha}} \sum_{\substack{g \in B_{p+1} \\ S(\alpha) \leq S(g) \restriction_p \\ p+1 \in V(S(g))}} prob(g) \cdot Cont(\beta | g)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{f\in B_{p+1}\\ S(\alpha)\leq S(f)\upharpoonright_p\\ p+1\notin V(S(f))}} \sum_{\substack{\beta\in SD_{p+1}\\ \beta\vDash\alpha}} prob(f)\cdot Cont(\beta|f)\\
&\quad+ \sum_{\substack{g\in B_{p+1}\\ S(\alpha)\leq S(g)\upharpoonright_p\\ p+1\in V(S(g))}} \sum_{\substack{\beta\in SD_{p+1}\\ \beta\vDash\alpha}} prob(g)\cdot Cont(\beta|g)\\
&= \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} \sum_{\substack{d\in B_1\\ d\in c}} prob(cd)\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))},1@t)}{N(\emptyset,1@t)} \text{ using } (*)\\
&\quad+ \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} \sum_{\substack{e\in B_1\\ e\notin c}} prob(ce) \sum_{\substack{\beta\in SD_{p+1}\\ \beta\vDash\alpha}} Cont(\beta|ce)\\
&= \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} \sum_{\substack{d\in B_1\\ d\in c}} prob(cd)\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))},1@t))}{N(\emptyset,1@t)}\\
&\quad+ \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} \sum_{\substack{e\in B_1\\ e\notin c}} prob(ce) \sum_{\substack{\beta\in SD_{p+1}\\ \beta\vDash\alpha\\ S(\beta)\leq S(ce)}} \frac{N(\overrightarrow{\mathcal{S}(\beta/S(ce))},1@t)}{N(\emptyset,1@t)}\\
&= \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} (\sum_{\substack{d\in B_1\\ d\in c}} prob(cd)\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))},1@t)}{N(\emptyset,1@t)}\\
&\quad+ \sum_{\substack{e\in B_1\\ e\notin c}} prob(ce) \sum_{\substack{\beta\in SD_{p+1}\\ \beta\vDash\alpha\\ S(\beta)\leq S(ce)}} \frac{N(\overrightarrow{\mathcal{S}(\beta/S(ce))},1@t)}{N(\emptyset,1@t)})\\
&= \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} [\sum_{\substack{d\in B_1\\ d\in c}} prob(c)prob(d)\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))},1@t)}{N(\emptyset,1@t)}\\
&\quad+ \sum_{\substack{e\in B_1\\ e\notin c}} prob(c)prob(e)\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))},1@t)}{N(\emptyset,1@t)}]\\
&= \sum_{\substack{c\in B_p\\ S(\alpha)\leq S(c)}} prob(c)Cont(\alpha|c)(\sum_{\substack{d\in B_1\\ d\in c}} prob(d) + \sum_{\substack{e\in B_1\\ e\notin c}} prob(e))\\
&= v^{\vec{p}}(\alpha)\ .
\end{aligned}
$$

□

We now give a small technical observation.

Lemma 17. For $q \in \mathbb{N}$ let $SubB_q \subset B_q$ such that $\lim_{q\to\infty} \sum_{c\in SubB_q} prob(c) = 0$, then

$$\lim_{q\to\infty} \sum_{\alpha\in SD_q} \sum_{c\in SubB_q} prob(c)Cont(\alpha|c) = 0 \ .$$

Proof. For every fixed $c \in B_q$ we have

$$\begin{aligned}
\sum_{\alpha\in SD_q} Cont(\alpha|c) &= \sum_{\substack{\alpha\in SD_q \\ S(\alpha)\le S(c)}} Cont(\alpha|c) \\
&= \sum_{\beta\in SD_{|S(c)|}} \frac{N(\overrightarrow{\mathcal{S}(\beta)}, 1@t)}{N(\emptyset, 1@t)} \\
&= \frac{1}{N(\emptyset, 1@t)} \sum_{\vec{x}\in X_{|S(c)|}} \sum_{\substack{\beta\in SD_{|S(c)|} \\ \overrightarrow{\mathcal{S}(\beta)}=\vec{x}}} N(\vec{x}, 1@t) \\
&= \frac{1}{N(\emptyset, 1@t)} \sum_{\vec{x}\in X_{|S(c)|}} N(\emptyset, \vec{x}) \cdot N(\vec{x}, 1@t) \\
&= 1 \ .
\end{aligned}$$

This explains the name contribution factor, as the sum over all contributions equals one. And so

$$\begin{aligned}
\sum_{\alpha\in SD_q} \sum_{c\in SubB_q} Cont(\alpha|c)prob(c) &= \sum_{c\in SubB_q} prob(c) \sum_{\alpha\in SD_q} Cont(\alpha|c) \\
&= \sum_{c\in SubB_q} prob(c) \ .
\end{aligned}$$

Now sending q to infinity completes the proof. □

Lemma 18. Let $\alpha, \beta \in SD_p$ such that $S(\alpha) = S(\beta)$ and $E \in ER_p$. If $S(\alpha) \le E$ then $S(\alpha/E) = S(\beta/E)$.

Proof. Note that $vip_{S(\alpha)}(j) = vip_{S(\alpha)}(v_E(j))$ since $S(\alpha)j(v_E(j))$. Let $1 \le i, k \le p$ be in $V(E)$. We have to show that $S(\alpha/E)(vip_E(i))(vip_E(k))$ if and only if $S(\beta/E)(vip_E(i))(vip_E(k))$.
Suppose α decides $R\vec{a}$ that is the constants in $\vec{a}$ are all in $\{a_1, \ldots, a_p\}$. Since $S(\alpha) \le E$

we have $\alpha \vDash R\vec{a}$ if and only if $\alpha/E \models Rvip_E(\vec{a})$. Hence

$$\begin{aligned}
&S(\alpha/E)(vip_E(i))(vip_E(k)) \\
\iff\ & \alpha/E \vDash \bigwedge_{R\in L} \bigwedge_{\substack{\vec{a},\vec{a}'\\ |\vec{a}|+|\vec{a}'|=arity(R)-1}} \\
& \qquad \big(Rvip_E(\vec{a})a_{vip_E(i)}vip_E(\vec{a}') \leftrightarrow Rvip_E(\vec{a})a_{vip_E(k)}vip_E(\vec{a}')\big) \\
\iff\ & \alpha \vDash \bigwedge_{R\in L} \bigwedge_{\substack{\vec{a},\vec{a}'\\ |\vec{a}|+|\vec{a}'|=arity(R)-1}} (R\vec{a}a_i\vec{a}' \leftrightarrow R\vec{a}a_k\vec{a}') \\
\iff\ & \beta \vDash \bigwedge_{R\in L} \bigwedge_{\substack{\vec{a},\vec{a}'\\ |\vec{a}|+|\vec{a}'|=arity(R)-1}} (R\vec{a}a_i\vec{a}' \leftrightarrow R\vec{a}a_k\vec{a}') \\
\iff\ & \beta/E \vDash \bigwedge_{R\in L} \bigwedge_{\substack{\vec{a},\vec{a}'\\ |\vec{a}|+|\vec{a}'|=arity(R)-1}} \\
& \qquad \big(Rvip_E(\vec{a})a_{vip_E(i)}vip_E(\vec{a}') \leftrightarrow Rvip_E(\vec{a})a_{vip_E(k)}vip_E(\vec{a}')\big) \\
\iff\ & S(\beta/E)(vip_E(i))(vip_E(k)) \ .
\end{aligned}$$

Where the third line is logically equivalent to the fourth because of $S(\alpha) = S(\beta)$. □

Definition 38. Let $c \in B_p$ and $i, k \leq p$. Then denote by $c(i \mapsto k, k \mapsto i)$ the result of swapping the $i-th$ with the $k-$th ball in c. Please note that $S(c(i \mapsto k, k \mapsto i)) = S(c)(i \mapsto k, k \mapsto i)$.

Theorem 8. *All the $v^{\vec{p}}$ satisfy* ***(SX)****.*

Proof. We first show that they satisfy **(CX)**.
Using the argument given after the proof of Lemma 6 we only need to prove the probability of state descriptions is invariant under transposing two constants.
For all $\alpha \in SD_p$ and all $1 \leq i < k \leq p$ we have that $S(\alpha) \leq S(c)$ if and only if $S(\alpha(a_k \mapsto a_i, a_i \mapsto a_k)) \leq S(c(k \mapsto i, i \mapsto k))$ and clearly $prob(c) = prob(c(k \mapsto i, i \mapsto k))$. Furthermore we have that the spectrum of $\alpha/S(c)$ equals the spectrum of $\alpha(a_k \mapsto a_i, a_i \mapsto a_k)/S(c(k \mapsto i, i \mapsto k))$ for $S(\alpha) \leq S(c)$.
Hence $v^{\vec{p}}(\alpha) = v^{\vec{p}}(\alpha(a_k \mapsto a_i, a_i \mapsto a_k))$.
Since we now know that the $v^{\vec{p}}$ satisfy the principle of constant exchangeability it only remains to show that for different $\alpha, \beta \in SD_p$ such that $S(\alpha) = S(\beta)$ that $v^{\vec{p}}(\alpha) = v^{\vec{p}}(\beta)$.
Since $S(\alpha) = S(\beta)$ we have that for every $c \in B_p$ that $S(\alpha) \leq S(c)$ if and only if

$S(\beta) \leq S(c)$. Furthermore by the above lemma $S(\alpha/S(c)) = S(\beta/S(c))$ and hence $\alpha/S(c)$ has the same spectrum as $\beta/S(c)$. Hence $v^{\vec{p}}(\alpha) = v^{\vec{p}}(\beta)$. □

Definition 39. Let C be the set of all draws that contain all t colors.

Lemma 19. The $v^{\vec{p}}$ are $t-$heterogeneous.

Proof. Clearly the $v^{\vec{p}}$ are $\leq t-$heterogeneous since we only have t colors in the urn. $c \in B_q \cap C$ and $Cont(\alpha|c) > 0$ imply that $\alpha \in SD_q^t$ since when we eventually draw the last remaining color we have to choose an extension with t classes according to equation (7.2). We observe that

$$\lim_{q\to\infty} \sum_{c\in B_q\setminus C} prob(c) \leq \lim_{q\to\infty} \sum_{i=1}^{t} (1-p_i)^q \leq \lim_{q\to\infty} t(1-p_t)^q = 0$$

since $p_{t-i} \geq p_t$ for $1 \leq i \leq t-1$. So

$$\lim_{p\to\infty} \sum_{c\in B_p\cap C} prob(c) = 1 \quad . \tag{7.3}$$

We find for all $q \geq t$ using $|\{\gamma \in SD_q | S(\gamma) = E\}| = N(\emptyset, 1@|E|)$

$$\begin{aligned}
\sum_{\gamma\in SD_q^t} v^{\vec{p}}(\gamma) &= \sum_{\gamma\in SD_q^t} \sum_{\substack{c\in B_q\\ S(\gamma)\leq S(c)}} prob(c) \frac{N(\overrightarrow{\mathcal{S}(\gamma/S(c))}, 1@t)}{N(\emptyset, 1@t)} \\
&= \sum_{\gamma\in SD_q^t} \sum_{\substack{c\in B_q\cap C\\ S(\gamma)=S(c)}} \frac{prob(c)}{N(\emptyset, 1@t)} \\
&= \sum_{E\in ER_q^t} \sum_{\substack{\gamma\in SD_q^t\\ S(\gamma)=E}} \sum_{\substack{c\in B_q\cap C\\ E=S(c)}} \frac{prob(c)}{N(\emptyset, 1@t)} \\
&= \sum_{E\in ER_q^t} \sum_{\substack{c\in B_q\cap C\\ E=S(c)}} prob(c) \\
&= \sum_{c\in B_q\cap C} prob(c) \quad .
\end{aligned}$$

Sending q to infinity shows that $v^{\vec{p}}$ is $t-$heterogeneous. □

Definition 40. Let

$$\mathbb{H}_t := \{\langle p_1, \dots, p_t\rangle \in \mathbb{R}^t \mid p_i \geq p_k > 0 \text{ for all } 1 \leq i < k \leq t \text{ and } \sum_{i=1}^{t} p_i = 1\}.$$

We endow $\mathbb{H}_t$ with the restriction of the standard topology on $\mathbb{R}^t$.

Definition 41. Let $n, m \in \mathbb{N}$ then put

$$Inj(n,m) := \{f : \{1,\dots,n\} \to \{1,\dots,m\} \mid f \text{ is injective}\} \ . \tag{7.4}$$

The following was first mentioned in [25] for binary languages and later proved in [18] for polyadic languages. Here we give a direct less involved proof. The advantage of the proof in [18] is that in can be adapted to prove the homogeneous representation theorem 11 on page 73. Unfortunately an adaption to the homogeneous setting of the here presented proof is not obvious as we here use theorem 3 on page 47 and theorem 4 on page 50. Since there are no homogeneous functions on a purely unary language we can't apply theorem 3 nor theorem 4 in the homogeneous case.
We remind the reader of our convention that $0! = 1$.

Theorem 9. *A de Finetti Representation Theorem* [2] *- The* $t-$*Heterogeneous Case*
Let w *be a probability function on* L*. Then* w *is* $t-$*heterogeneous if and only if there exists a measure* [3] μ *on* $\mathbb{H}_t$ *such that for all* $\psi \in QFSL$

$$w(\psi) = \int_{\mathbb{H}_t} v^{\vec{p}}(\psi) \, d\mu(\vec{p}) \ . \tag{7.5}$$

Proof. If w is given in such a form then clearly it is $t-$heterogeneous as it is a generalized convex combination of $t-$heterogeneous functions.
Now for the other direction assume that n is minimal such that $t \leq 2^n$. We define a $t-$heterogeneous function on the purely unary language L' with n unary relation symbols. As above we define the probability function as the result of drawing balls from an urn and then we take the thereby induced probability function on L'. The so defined probability function will be called $v'^{\vec{p}}$ to indicate that it is on L'.
In case of a purely unary language constants that look the same over a state description

[2]The name is due to Bruno de Finetti who first proved such a theorem for purely unary languages, see [7].

[3]We will always assume that our *measures* μ are *probability measures* over the space Z considered. That is $\int_Z d\mu = 1$. This notion of measure hence generalizes the idea of a convex combination.

will look the same over every extension of this state description. Again we want to concentrate in the limit all measure on state descriptions with spectrum of length t. Hence whenever we draw a new color we have to choose an extension that does separate the new constant from the rest.

Supposing we have already seen s colors in our draw then there are $2^n - s$ possible choices left to distinguish the new constant for this extension since there are 2^n atoms over L'. So the contribution-factor is

$$\frac{N(1@s+1, 1@t)}{N(1@s, 1@t)} = \frac{(2^n - s - 1)!}{(2^n - s)!} = \frac{1}{2^n - s}$$

when we draw a new color. This phenomenon is responsible for the $(2^n - R)!/2^n!$ factor in the definition of $v'^{\vec{p}}$ below.

We let for $\alpha \in SD_q^R(L')$ with spectrum $\vec{x} \in X_q^R$

$$v'^{\vec{p}}(\alpha) := \frac{(2^n - R)!}{2^n!} \sum_{\substack{c \in B_q \\ S(\alpha)=S(c)}} prob(c) = \frac{(2^n - R)!}{2^n!} \sum_{f \in Inj(R,t)} \prod_{i=1}^{R} p_{f(i)}^{x_i} \quad . \tag{7.6}$$

$v'^{\vec{p}}$ is a probability function on L' because of

$$\sum_{\alpha \in SD_1(L')} v'^{\vec{p}}(\alpha) = \sum_{\alpha \in SD_1(L')} \frac{1}{2^n} \sum_{f \in Inj(1,t)} p_{f(1)} = \frac{2^n}{2^n} \sum_{i=1}^{t} p_i = 1$$

and because for $\alpha \in SD_q^R(L')$ with spectrum $\vec{x} \in X_q^R$ we have

$$\begin{aligned}
\sum_{\substack{\beta \in SD_{q+1}(L') \\ \beta \vDash \alpha}} v'^{\vec{p}}(\beta) = & \sum_{\substack{\beta \in SD_{q+1}^{R+1}(L') \\ \beta \vDash \alpha}} v'^{\vec{p}}(\beta) + \sum_{\substack{\beta \in SD_{q+1}^{R}(L') \\ \beta \vDash \alpha}} v'^{\vec{p}}(\beta) \\
=&(2^n - R) \cdot \frac{(2^n - (R+1))!}{2^n!} \sum_{f \in Inj(R+1,t)} p_{f(R+1)} \prod_{i=1}^{R} p_{f(i)}^{x_i} \\
&+ \sum_{j=1}^{R} \frac{(2^n - R)!}{2^n!} \sum_{f \in Inj(R,t)} p_{f(j)} \prod_{i=1}^{R} p_{f(i)}^{x_i} \\
=&\frac{(2^n - R)!}{2^n!} \sum_{f \in Inj(R,t)} \Big(\sum_{\substack{1 \leq l \leq t \\ l \notin Im(f)}} p_l \Big) \cdot \prod_{i=1}^{R} p_{f(i)}^{x_i}
\end{aligned}$$

$$+\frac{(2^n-R)!}{2^n!}\sum_{f\in Inj(R,t)}(\sum_{j=1}^{R}p_{f(j)})\cdot\prod_{i=1}^{R}p_{f(i)}^{x_i}$$
$$=\frac{(2^n-R)!}{2^n!}\sum_{f\in Inj(R,t)}\prod_{i=1}^{R}p_{f(i)}^{x_i}$$
$$=v'^{\vec{p}}(\alpha)\ .$$

According to the representation theorem 4 on page 50 $v'^{\vec{p}}$ induces a $t-$heterogeneous probability function on our given polyadic language L which will be denoted by $w_{\vec{p}}$. Recall from equation 5.25 on page 51 that then

$$w_{\vec{p}}(\vec{x})=\lim_{q\to\infty}\sum_{\vec{y}\in X_q^t}d(t,L',L)N(\vec{x},\vec{y},L)v'^{\vec{p}}(\vec{y})\ .$$

As we have seen the right hand side defines a probability function and hence the limit exists.
Let v be the probability function on L' satisfying **(SX)** that is the representation as in theorem 3 on page 47 of the given w. De Finetti's celebrated theorem states that every probability function on a purely unary langauge that satisfies **(CX)** can be written as a certain integral (see [7]). By a generalization of Zabell to probability functions that also satisfy **(AX)** (see [30]) and since v is $t-$heterogeneous there is a measure μ on $\mathbb{H}_t$ such that

$$v(\vec{x})=\int_{\mathbb{H}_t}v'^{\vec{p}}(\vec{x})\ d\mu(\vec{p})\ .$$

We'll use the theorem of dominated convergence in the appendix (see page 151 Corollary 15). As the sequence $(w_q^{\vec{p}})_{q\in\mathbb{N}}$ we here take $\sum_{\vec{y}\in X_q^t}d(t,L',L)N(\vec{x},\vec{y},L)v'^{\vec{p}}(\vec{y})=:$ $w_q^{\vec{p}}(\vec{x})$. These are in general not probability functions but $\lim_{q\to\infty}w_q^{\vec{p}}$ is a probability function and for all q we have that $w_q^{\vec{p}}(\psi)\leq w_{q+1}^{\vec{p}}(\psi)$ for all $\psi\in QFSL$. So the domination condition of the theorem in the appendix is clearly satisfied.

Hence we can apply the corollary and get for this measure μ that

$$\begin{aligned}\int_{\mathbb{H}_t} w_{\vec{p}}(\vec{x})\, d\mu(\vec{p}) &= \int_{\mathbb{H}_t} \lim_{q\to\infty} \sum_{\vec{y}\in X_q^t} d(t,L',L)N(\vec{x},\vec{y},L)v'^{\vec{p}}(\vec{y})d\mu \\ &= \lim_{q\to\infty} \sum_{\vec{y}\in X_q^t} d(t,L',L)\cdot N(\vec{x},\vec{y},L) \int_{\mathbb{H}_t} v'^{\vec{p}}(\vec{y})\, d\mu(\vec{p}) \\ &= \lim_{q\to\infty} \sum_{\vec{y}\in X_q^t} d(t,L',L)\cdot N(\vec{x},\vec{y},L)\, v(\vec{y}) \\ &= w(\vec{x}) \ .\end{aligned}$$

Let $v^{\vec{p}}$ be the $t-$heterogeneous functions on L as defined in equation (7.2) on page 61 for a given $\vec{p} \in \mathbb{H}_t$. We will now show that $v^{\vec{p}} = w_{\vec{p}}$. Due to the $t-$heterogeneity of $v^{\vec{p}}$ and $w_{\vec{p}}$ it is enough to prove that $v^{\vec{p}}(\vec{x}) = w_{\vec{p}}(\vec{x})$ for all $\vec{x} \in X^t$.
Since $v'^{\vec{p}}$ is $t-$heterogeneous we have that $w_{\vec{p}}(\vec{x}) = d(t,L',L)\cdot v'^{\vec{p}}(\vec{x})$ for $\vec{x} \in X^t$. Now pick a $\beta \in SD_q^t(L')$ with spectrum $\vec{x}$. Then

$$w_{\vec{p}}(\vec{x}) = d(t,L',L)\cdot v'^{\vec{p}}(\beta) = d(t,L',L)\cdot \frac{(2^n-t)!}{2^n!} \sum_{f\in Inj(t,t)} \prod_{i=1}^{t} p_{f(i)}^{x_i} \ .$$

Let $\alpha \in SD_q^t(L)$ have spectrum $\vec{x}$. Then since $N(\emptyset, 1@t, L') = 2^n!/(2^n-t)!$

$$\begin{aligned}v^{\vec{p}}(\vec{x}) &= \left(\sum_{\sigma\in S^t} \prod_{i=1}^{t} p_{\sigma(i)}^{x_i}\right)\cdot N(\emptyset,1@t,L)^{-1} \\ &= \left(\sum_{\sigma\in S^t} \prod_{i=1}^{t} p_{\sigma(i)}^{x_i}\right)\frac{(2^n-t)!}{(2^n!)}\cdot d(t,L',L) \\ &= d(t,L',L)\frac{(2^n-t)!}{(2^n!)}\cdot\left(\sum_{f\in Inj(t,t)} \prod_{i=1}^{t} p_{f(i)}^{x_i}\right) = w_{\vec{p}}(\vec{x}) \ .\end{aligned}$$

Hence $v^{\vec{p}}$ and $w_{\vec{p}}$ are the same on X^t and so $v^{\vec{p}} = w_{\vec{p}}$. □

7.2 The Homogeneous Case

Most of the following can be found in [18]. Here we give some intuition and fill out some details. The proof of the main theorem 11 however, will be omitted here. The interested reader is referred to [18].

Generally speaking the homogeneous case is not too different from the heterogeneous case. Here we require that our urn contains infinitely many colors or that it contains a ball that is black with non-zero probability of being drawn.[4]

Suppose we are given an urn containing colored balls with pairwise different colors $c_0, c_1, \ldots$. Every color c_i comes with a fixed probability p_i of being drawn such that $\sum_{i=0}^{\infty} p_i = 1$. We say that the c_0 is black. [5]
Furthermore we assume that for $1 \leq i < k$ we have $p_i \geq p_k \geq 0$ and $p_0 \geq 0$. As indicated above we either require that $p_0 > 0$ or all other $p_i > 0$ or both.
Let B_p be as before and define an equivalence relation $S(c)$ for $c \in B_p$ on p by letting $S(c)ik$ if and only if the $i-$th and the $k-$th ball in c have the same color and are not black. If the $i-$th ball in c is black then i is a singleton in $S(c)$.
So if there's no black ball in c then this definition of $S(c)$ and the previous one agree.
Given such an urn we will inductively construct a homogeneous probability function $u^{\vec{p}}$.

First step: Pick one ball from the urn and choose with the uniform distribution an $\alpha^1 \in SD_1$.
Inductive step from $p-1$ **to** p: Assume we have so far constructed an $\alpha^{p-1} \in SD_{p-1}$ according to a draw in B_{p-1}. Pick another ball from the urn and let c denote the draw of the first p balls. That is we have our first $p-1$ balls and then we draw the $p-$th ball. If the $p-$th ball has the same color as the $i-th$ ball and is not black then uniquely extend α^{p-1} to α^p such that $S(\alpha^p)pi$ holds.
If the $p-$th ball has a color not among the previously chosen colors or is black then uniformly randomly chose an α^p extending α^{p-1} that satisfies that if the $i-$th and the $j-$th ball have the same color and are not black then $S(\alpha^p)ij$. That is one uniformly randomly picks an element in $\{\alpha^p \in SD_p | \alpha^p \vDash \alpha^{p-1},\ S(\alpha^p) \leq S(c)\}$.
The probability that for a given draw c we will choose a given state description $\alpha \in SD_p$ will be denoted by $Contr(\alpha|c)$.
Note that if $S(c)$ is not a refinement of $S(\alpha)$ then $Contr(\alpha|c) = 0$, since two balls of the same color in c lead to constants that are indistinguishable.
Now let

[4] To make it clear we also allow that there are infinitely many colors *and* a black ball.

[5] We make the convention that black is **not** a color. The reason for this convention will come apparent later on as the black ball plays a special role.

$$u^{\vec{p}}_L(\alpha) := \sum_{\substack{c\in B_p \\ S(\alpha)\leq S(c)}} prob(c)Contr(\alpha|c) = \sum_{\substack{c\in B_p \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}(L)|} \ . \tag{7.7}$$

The subscript L indicating the language will be dropped when there can't be any confusion.

Alternatively we could have introduced the $u^{\vec{p}}$ such that for a given draw $c \in B_p$ we uniformly chose an $\alpha \in SD_{|S(c)|}$, then blow it up to $\alpha \cdot S(c)$ and then consider the induced probability function.

We feel that the former way is more graphic. The latter however is better for calculations.

Theorem 10. *All the $u^{\vec{p}}$ satisfy **(SX)** and **(REG)**.*

Proof. We first show that they satisfy **(CX)**.

Using the same argument as in the proof of theorem 8 on page 65 we only need to prove that the probability of state descriptions is invariant under transposing constants. For all $\alpha \in SD_p$, $c \in B_p$ and all $1 \leq i < k \leq p$ we have that $S(\alpha) \leq S(c)$ if and only if $S(\alpha(a_i \mapsto a_k, a_k \mapsto a_i)) \leq S(c(i \mapsto k, k \mapsto i))$.

Hence $u^{\vec{p}}(\alpha) = u^{\vec{p}}(\alpha(a_i \mapsto a_k, a_k \mapsto a_i))$ by equation (7.7).

Since we now know that the $u^{\vec{p}}$ satisfy the principle of constant exchangeability it only remains to show that for different $\alpha, \beta \in SD_p$ such that $S(\alpha) = S(\beta)$ that $u^{\vec{p}}(\alpha) = u^{\vec{p}}(\beta)$.

This however follows directly from equation (7.7) since the value of $u^{\vec{p}}$ only depends on $S(\alpha)$ and not on the state description itself.

Since we have either infinitely many colors or a black ball in our urn we can for all p always find a $c \in B_p$ such that $S(c)$ consists only of singletons. For such a c and every $\gamma \in SD_p$ we have $Contr(\gamma|c) = 1/|SD_p|$ and hence $u^{\vec{p}}(\gamma) > 0$.

It follows from Lemma 3 on page 17 that for $\psi \in QFSL$ it has to hold that $u^{\vec{p}}(\psi) = \sum_{\substack{\beta\in SD_q \\ \beta\vDash\psi}} u^{\vec{p}}(\beta)$ for some suitable (i.e. big enough) q. If ψ is satisfiable then the sum is over at least one state description and hence we have $u^{\vec{p}}(\psi) > 0$.

So the $u^{\vec{p}}$ satisfy **(REG)**. □

Definition 42. For $i \geq 0$ and $x_i \in \mathbb{R}$ with $x_i \geq 0$ and $t \geq 1$ let

$$\mathbb{B} := \{\langle x_0, x_1, \dots \rangle \mid x_1 \geq x_2 \geq x_3 \dots, \sum_{i=0}^{\infty} x_i = 1 \} \tag{7.8}$$

$$\mathbb{H}_0 := \{\langle x_1, x_2, \dots \rangle \mid x_1 \geq x_2 \geq x_3 \dots \text{ and for all } k > 0 : \sum_{i=1}^{k} x_i < 1\} \tag{7.9}$$

$$\mathbb{H}_t := \{\langle x_1, x_2, \dots, x_t \rangle \mid x_1 \geq x_2 \geq \dots \geq x_t > 0 : \sum_{i=1}^{t} x_i = 1\} \tag{7.10}$$

$$\mathbb{H}_\infty := \mathbb{H}_0 \cup \bigcup_{t=1}^{\infty} \mathbb{H}_t \ . \tag{7.11}$$

Define a map $h : \mathbb{H}_0 \to \mathbb{B}$ via $h(x_1, x_2 \dots) := \langle 1 - \sum_{i=1}^{\infty} x_i, x_1, x_2, \dots \rangle$.
We endow $\mathbb{B}, \mathbb{H}_\infty$ and $\mathbb{H}_0$ with the standard weak topology inherited from $[0,1]^\infty$.

Theorem 11. *A probability function w on a not purely unary L is homogeneous if and only if there exists a measure μ_0 on $\mathbb{H}_0$ such that*

$$w = \int_{\mathbb{H}_0} u^{h(\vec{p})} d\mu_0(\vec{p}) \ . \tag{7.12}$$

As mentioned above, a proof can be found in [18].

Corollary 4. Let w satisfy **(SX)** on a not purely unary language L. Then for $0 \leq i$ there are measures μ_i on $\mathbb{H}_i$ and unique $\eta_i \geq 0$[6] such that[7]

$$w = \eta_0 \int_{\mathbb{H}_0} u^{h(\vec{p})} d\mu_0 + \sum_{i=1}^{\infty} \eta_i \int_{\mathbb{H}_i} v^{\vec{p}} d\mu_i \ . \tag{7.13}$$

Let $s^{\vec{p}} := u^{h(\vec{p})}$ for $\vec{p} \in \mathbb{H}_0$ and $s^{\vec{p}} := v^{\vec{p}}$ if $\vec{p} \in \mathbb{H}_t$. *If w satisfies **(SX)** then there is a measure ν on $\mathbb{H}_\infty$ such that*

$$w = \int_{\mathbb{H}_\infty} s^{\vec{p}} d\nu \ . \tag{7.14}$$

If L is purely unary then ν concentrates all measure on $\bigcup_{i=1}^{2^{m_1}} \mathbb{H}_i$.

It's now worth pointing out that if $w(\vec{x}) = 0$ for some $\vec{x} \in X^t$ then $\eta_{t+i} = \eta_0 = 0$ for all $i \geq 0$.
$\eta_{t+i} = 0$ for the following reason. Suppose w' is $t+i-$heterogeneous. Then consider

[6] Where the η are the coefficients of the $\eta-$representation of w defined as in theorem 5 on page 52.
[7] The function $\int_{\mathbb{H}_i} v^{\vec{p}} d\mu_i$ is unique whenever $\eta_i > 0$.

the draw $c = \langle c_1, c_2, \ldots, c_t \rangle$ with respect to some $\vec{p} \in \mathbb{H}_{t+i}$. Clearly $S(c) \in ER_t^t$ and so $v^{\vec{p}}(\langle 1@t \rangle > 0$ for all $\vec{p} \in \mathbb{H}_{t+i}$. But now it is easy to find a draw $d \in B^{\vec{p}}$ extending c such that $S(d) = S(\alpha)$ for some $\alpha \in SD^t$ with spectrum $\vec{x}$. Hence $v^{\vec{p}}(\vec{x}) > 0$ for all $\vec{p} \in \mathbb{H}_{t+i}$ and so $w'(\vec{x}) > 0$. Hence w can't contain a $t+i-$heterogeneous part.
$\eta_0 = 0$ follows from the fact every homogeneous probability function satisfies **(REG)** since all the $u^{\vec{p}}$ do.

Definition 43. Let w_∞ be the unique probability function that satisfies for all p and all $\alpha \in SD_p(L)$ that $w_\infty(\alpha) = \frac{1}{|SD_p(L)|}$. That is w_∞ treats all atomic formulae as stochastically independent. We call w_∞ the *completely independent function*.

Remark 6. There are probability functions satisfying **(SX)** that are neither homogeneous nor $\leq t-$heterogeneous for any t.

Proof. Pick $\eta_i > 0$ for $i > 0$ such that $\sum_{i=1}^{\infty} \eta_i = 1$ and for all i an $i-$heterogeneous distribution w_i.[8] Now consider $w := \sum_{i=1}^{\infty} \eta_i w_i$. Clearly this probability function satisfying **(REG)** and **(SX)**. Furthermore it's not homogeneous since

$$\lim_{p \to \infty} \sum_{\alpha \in SD_p^t} w(\alpha) = \eta_t \quad .$$

Another example is a convex combination of η_0 and η_t of a homogeneous w_0 and a $t-$heterogeneous $w^{[t]}$ then $w := \eta_0 w_0 + (1 - \eta_t) w^{[t]}$ does the trick for $\eta_0 \neq 0 \neq \eta_t$. The simplest such example is $w := \frac{w^{[1]} + w_\infty}{2}$. □

We will now give a small technical lemma that is the analogue of Lemma 17 on page 64.

Lemma 20. For $q \in \mathbb{N}$ let $SubB_q \subset B_q$ such that $\lim_{q \to \infty} \sum_{c \in SubB_q} prob(c) = 0$, then

$$\lim_{q \to \infty} \sum_{\alpha \in SD_q} \sum_{c \in SubB_q} prob(c) Contr(\alpha|c) = 0 \quad .$$

Proof. For every fixed $c \in B_q$ we have

$$\sum_{\alpha \in SD_q} Contr(\alpha|c) = \sum_{\substack{\alpha \in SD_q \\ S(\alpha) \leq S(c)}} Contr(\alpha|c) = \sum_{\beta \in SD_{|S(c)|}} \frac{1}{|SD_{|S(c)|}|} = 1 \quad .$$

[8] If L is purely unary then consider a convex combination of $\eta_1, \ldots, \eta_{2^{m_1}}$ of some $i-$heterogeneous $w^{[i]}$.

And so

$$\begin{aligned}\sum_{\alpha \in SD_q} \sum_{c \in SubB_q} Contr(\alpha|c)prob(c) &= \sum_{c \in SubB_q} prob(c) \sum_{\alpha \in SD_q} Cont(\alpha|c) \\ &= \sum_{c \in SubB_q} prob(c) \ .\end{aligned}$$

Now sending q to infinity completes the proof. □

Chapter 8

Language Invariance

In this section we only consider probability functions that satisfy **(SX)**. The main theorem in this chapter theorem 12 was proved in [16]. The sufficiency condition already appeared in [15].

Definition 44. Let L, L' be languages in our sense (see definitions 6 and 7). We say that L' is an *extension* of L if and only if for all $1 \leq d \leq r_0(L)$ we have that $L_d \subset L'_d$. In that case we say that L is a *sublanguage* of L'. So for every arity there are at least as many relation symbols in L'_d as there are in L_d.

Definition 45. A *language invariant family* of probability functions $w_{\mathcal{L}}$ is a family of probability functions $w_{\mathcal{L}}$, such that for any language L (in our sense), each satisfying **(SX)** such that if L is a sublanguage of L' and $\varphi \in SL$ then $w_L(\varphi) = w_{L'}(\varphi)$.
A probability function w is said to be *language invariant* if it is a member of a language invariant family.

As it turns out there's a rather close connection between the $u^{\vec{p}}$ and the notion of language invariance.
Previously we defined probability functions $u^{\vec{p}}$ with $h^{-1}(\vec{p}) \in \mathbb{H}_0$. Now we will define probability functions for $\vec{p} \in \mathbb{B}$ still denoted by $u^{\vec{p}}$. If $\vec{p}$ is such that $h^{-1}(\vec{p}) \in \mathbb{H}_0$ then both definitions agree. We put as in equation (7.7)

$$u_L^{\vec{p}}(\alpha) := \sum_{\substack{c \in B_p \\ S(\alpha) \leq S(c)}} prob(c) Contr(\alpha|c) = \sum_{\substack{c \in B_p \\ S(\alpha) \leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}(L)|} \ . \tag{8.1}$$

The same arguments as above show that $u_L^{\vec{p}}$ is also a probability function if $\vec{p} \notin Im(h)$. The main difference is that if $\vec{p}$ is such that $p_0 = 0$ and such that there is a k such that

$p_k = 0$ then $Contr(\alpha|c) = 0$ for state descriptions with too long a spectrum for every draw c. Hence for such $\vec{p}$ $u^{\vec{p}}$ can't be homogeneous.

Theorem 12. *Let w be a probability function on L satisfying **(SX)**.*
Then a probability function w on L is language invariant if and only if there's a measure μ on $\mathbb{B}$ such that

$$w = \int_{\mathbb{B}} u_L^{\vec{p}} \, d\mu(\vec{p}) \quad . \tag{8.2}$$

If $L \subset L'$ and L is not purely unary then the extension $w_{L'}$ of w to L' is unique and

$$w_{L'} = \int_{\mathbb{B}} u_{L'}^{\vec{p}} \, d\mu(\vec{p}) \tag{8.3}$$

for the same measure μ from above. Furthermore a language invariant family $w_{\mathcal{L}}$ is a given by a measure μ such that for all languages L

$$w_L = \int_{\mathbb{B}} u_L^{\vec{p}} \, d\mu \quad . \tag{8.4}$$

For a proof see [16].

Lemma 21. Let $w_{\mathcal{L}}, v_{\mathcal{L}}$ be two language invariant families such that there is a language L and a spectrum $\vec{x} \in X_p^t$ such that $w_L(\vec{x}) \neq v_L(\vec{x})$.
Then there is purely unary language L^1 such that $w_{L^1} \neq v_{L^1}$.

Proof. Let L^1 be a purely unary language containing p relation symbols such that $L^1 \cap L = \emptyset$. Let L' be the language that contains the relation symbols in L and L^1. We write $L' = L \cup L^1$. Note that for $\alpha \in SD_p^t(L)$ with spectrum $\vec{x}$ we have

$$\sum_{\beta \in SD_p(L^1)} w_{L'}(\alpha \wedge \beta) = w_{L'}(\alpha) = w_L(\alpha) \neq v_L(\alpha) = v_{L'}(\alpha) = \sum_{\beta \in SD_p(L^1)} v_{L'}(\alpha \wedge \beta).$$

Hence there is a $\gamma \in SD_p(L')$ such that $w_{L'}(\gamma) \neq v_{L'}(\gamma)$. Now choose the minimal q such that there is a $\vec{y} \in X_q$ such that $w_{L'}(\vec{y}) \neq v_{L'}(\vec{y})$. Let s be such that for all $\vec{z} \in X_q^{>s}$ we have $w_{L'}(\vec{z}) = v_{L'}(\vec{z})$. If for example $w_{L'}(1@q) \neq v_{L'}(1@q)$ then we pick s to be q.
Letting $\alpha \in SD_q^s(L^1)$ with spectrum $\vec{r} \in X_q^s$ such that $w_{L'}(\vec{r}) \neq v_{L'}(\vec{r})$ we find

$$w_{L^1}(\alpha) = \sum_{\beta \in SD_q(L)} w_{L'}(\alpha \wedge \beta)) \tag{8.5}$$

$$= \sum_{\substack{\beta \in SD_q(L)\\ S(\beta)\leq S(\alpha)}} w_{L'}(\alpha\wedge\beta) + \sum_{\substack{\gamma \in SD_q(L)\\ \text{not } S(\gamma)\leq S(\alpha)}} w_{L'}(\alpha\wedge\gamma) \tag{8.6}$$

$$= \sum_{\substack{\beta \in SD_q(L)\\ S(\beta)\leq S(\alpha)}} w_{L'}(\vec{r}) + \sum_{\substack{\gamma \in SD_q(L)\\ \text{not } S(\gamma)\leq S(\alpha)}} w_{L'}(\alpha\wedge\gamma) \tag{8.7}$$

$$\neq \sum_{\substack{\beta \in SD_q(L)\\ S(\beta)\leq S(\alpha)}} v_{L'}(\vec{r}) + \sum_{\substack{\gamma \in SD_q(L)\\ \text{not } S(\gamma)\leq S(\alpha)}} w_{L'}(\alpha\wedge\gamma) \tag{8.8}$$

$$= \sum_{\substack{\beta \in SD_q(L)\\ S(\beta)\leq S(\alpha)}} v_{L'}(\alpha\wedge\beta) + \sum_{\substack{\gamma \in SD_q(L)\\ \text{not } S(\gamma)\leq S(\alpha)}} v_{L'}(\alpha\wedge\gamma) \tag{8.9}$$

$$= \sum_{\beta \in SD_q(L)} v_{L'}(\alpha\wedge\beta) \tag{8.10}$$

$$= v_{L^1}(\alpha) \ . \tag{8.11}$$

Where we used for (8.9) that if $S(\alpha)$ is not a refinement of $S(\gamma)$ then there are two constants that we can tell apart over $\alpha \wedge \gamma$ that look the same over α and hence the spectrum of $\alpha \wedge \gamma$ is longer than the spectrum of α. So the spectrum is in $X_q^{>s}$ and $v_{L'}$ and $w_{L'}$ agree by construction on $X_q^{>s}$. □

Corollary 5. Let $w_{\mathcal{L}}, v_{\mathcal{L}}$ be language invariant families such that $w_{L^1} = v_{L^1}$ for every purely unary language L^1. Then $w_L = v_L$ for every language L.

Lemma 22. Let $\vec{p} \in \mathbb{B}$ and let L be not purely unary. Then $u_L^{\vec{p}}$ can be approximated arbitrary closely on $SD_{\leq k}(L)$ by a sequence $(v_L^{\vec{q}_n})_{n\in\mathbb{N}}$ with $\vec{q}_n \in \mathbb{H}_n$.

Proof. First we show that if $p_0 > 0$ then we can approximate $u^{\vec{p}}$ arbitrary closely by a sequence $u^{\vec{p}_n}$ with $\vec{p}_n \in \mathbb{B}$ on SD_k such that all the $\vec{p}_n$ give the black ball probability zero of being drawn.

We put $\vec{p}_n := \langle 0, p_1, \ldots, p_r, \frac{p_0}{n}@n, p_{r+1}, p_{r+2}, \ldots\rangle$ where r [1] is such that $p_r > p_0/n \geq p_{r+1}$. We think of the newly inserted colors as *shades of grey* with the convention that none of balls originally in the urn was grey. Let $B_k^{\vec{p}_n}$ denote the set of draws of k balls with respect to $\vec{p}_n$.

Now consider an $\alpha \in SD_k$ and a $c \in B_k$ [2] with $S(c) \geq S(\alpha)$. If there is no black ball in c then we can think of c also as being a draw in $B_k^{\vec{p}_n}$. Hence such a draw contributes in $u^{\vec{p}}$ the same amount to the probability of α as in $u^{\vec{p}_n}$.

Now if c does contain a black ball we then consider the set C_n of draws $c_n \in B_k^{\vec{p}_n}$ that

[1] r depends on n

[2] If there's no superscript on B_k then we consider B_k to be with respect to the originally given $\vec{p}$.

have any grey ball at the place where a black ball was in c.[3] Increasing the number of shades of grey and replacing the black ball by a particular grey balls with probability p_0/n shows that the probability of $S(c_n) = S(c)$ converges to one as n approaches infinity, since with increasing n the probability of drawing the same shade of grey twice converges to zero. So the probability that a grey ball is responsible for a singleton in $S(c_n)$ converges to one. Hence for a fixed $c \in B_k$ we have

$$\lim_{n\to\infty} \sum_{\substack{c_n \in C_n \\ S(c_n)=S(c)}} Contr(\alpha|c_n)prob(c_n) = prob(c)Contr(\alpha|c) \ . \tag{8.12}$$

So we get for $\alpha \in SD_k$

$$\begin{aligned}
u^{\vec{p}}(\alpha) &= \sum_{\substack{c\in B_k \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} \\
&= \sum_{\substack{c\in B_k \\ c_0\notin c \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} + \sum_{\substack{c\in B_k \\ c_0\in c \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} \\
&= \sum_{\substack{c\in B_k \\ c_0\notin c \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} + \lim_{n\to\infty} \sum_{\substack{c\in B_k \\ c_0\in c \\ S(\alpha)\leq S(c)}} \sum_{\substack{c_n\in C_n \\ S(c_n)=S(c)}} \frac{prob(c_n)}{|SD_{|S(c_n)|}|} \\
&= \lim_{n\to\infty} u^{\vec{p}_n}(\alpha) \ .
\end{aligned}$$

Next we show that if we have infinitely many colors but no black in $\vec{p}$ then $u^{\vec{p}}$ can be approximated on SD_k by $(u^{\vec{q}_m})_{m\in\mathbb{N}}$ with $\vec{q}_m \in \mathbb{H}_{m+1}$, that is there are finitely many colors and no black ball are in our urn.

Let D_m be the set of draws of balls that contains only the first m colors with respect to $\vec{p}$ and let $\vec{q}_m := \langle p_1, \ldots, p_{r_m}, 1 - \sum_{i=1}^m p_i, p_{r_m+1}, \ldots, p_m\rangle$ where r_m is such that $p_{r_m} > 1 - \sum_{i=1}^m p_i \geq p_{r_m+1}$ and hence $\vec{q}_m \in \mathbb{H}_{m+1}$. Please note that r_m approaches infinity as m gets ever greater. We have for any fixed k

$$\lim_{m\to\infty} \sum_{c\in B_k\setminus D_m} prob(c) \leq \lim_{m\to\infty}((\sum_{r>m} p_r)^k) = (\lim_{m\to\infty}\sum_{r>m} p_r)^k = 0$$

[3] So C_n depends on c.

and hence

$$\lim_{m\to\infty} \sum_{\substack{c\in B_k^{\vec{q}_m}\\ c_{r_m+1}\in c}} prob(c) = 0 \ \ .$$

These observations now yield for a fixed $E \in ER_k$ that

$$\begin{aligned}
\lim_{m\to\infty} \sum_{\substack{c\in B_k\\ S(c)=E}} prob(c) &= \lim_{m\to\infty} \sum_{\substack{c\in B_k\\ S(c)=E\\ c\in D_m}} prob(c) + \lim_{m\to\infty} \sum_{\substack{c\in B_k\\ S(c)=E\\ c\notin D_m}} prob(c)\\
&= \lim_{m\to\infty} \sum_{\substack{c\in B_k\\ S(c)=E\\ c\in D_m}} prob(c)\\
&= \lim_{m\to\infty} \sum_{\substack{c\in B_k^{\vec{q}_m}\\ S(c)=E\\ c_{r_m+1}\notin c}} prob(c)\\
&= \lim_{m\to\infty} \sum_{\substack{c\in B_k^{\vec{q}_m}\\ S(c)=E\\ c_{r_m+1}\notin c}} prob(c) + \lim_{m\to\infty} \sum_{\substack{c\in B_k^{\vec{q}_m}\\ S(c)=E\\ c_{r_m+1}\in c}} prob(c)\\
&= \lim_{m\to\infty} \sum_{\substack{c\in B_k^{\vec{q}_m}\\ S(c)=E}} prob(c)
\end{aligned}$$

and so

$$\begin{aligned}
u^{\vec{p}}(\alpha) &= \sum_{\substack{c\in B_k\\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|}\\
&= \sum_{\substack{c\in B_k\\ c\in D_m\\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} + \sum_{\substack{c\in B_k\\ c\notin D_m\\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|}\\
&= \lim_{m\to\infty} \sum_{\substack{c\in B_k\\ c\in D_m\\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} + \lim_{m\to\infty} \sum_{\substack{c\in B_k\\ c\notin D_m\\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|}\\
&= \lim_{m\to\infty} \sum_{\substack{c\in B_k^{\vec{q}_m}\\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|}\\
&= \lim_{m\to\infty} u^{\vec{q}_m}(\alpha) \ \ .
\end{aligned}$$

Next we prove that every such $u^{\vec{q}_m}$ can be approximated by a sequence $(v^{\vec{n}_s})_{s\in\mathbb{N}}$ with $\vec{n}_s \in \mathbb{H}_{m+1+s}$.
Put $\vec{n}_s := \langle q_1, \ldots, q_m, q_{m+1} - \frac{q_{m+1}}{s}, \frac{q_{m+1}}{s^2}@s\rangle$ for $s \geq 2$. Then $\vec{n}_s \in \mathbb{H}_{m+1+s}$. Let $B_k^{\vec{n}_s}$ be the set of draws of k balls with respect to $\vec{n}_s$. In the limit as $s \to \infty$ the sum of probability of draws of k balls containing only the first $m+1$ colors goes to one. With other words the combined contribution of all draws of k balls containing any of the the last s colors is negligible for large s.
We prove in [18] page 14 that for $\vec{x} \in X_g$

$$\lim_{s\to\infty} N(\vec{x}, 1@s)/N(\emptyset, 1@s) = 1/|SD_g| \ . \tag{8.13}$$

The idea here is that for a huge s and a random $\alpha \in SD_s$ extending a $\beta \in SD_g$ with spectrum $\vec{x}$ it is *very* unlikely that $S(\alpha)ik$ for any fixed $1 \leq i < k \leq s$.
We denote by F_s all those draws in $B_k^{\vec{n}_s}$ that contain only the first most likely $m+1$ colors with respect to $\vec{n}_s$. For $\alpha \in SD_k$ we now find using (8.13) to obtain (8.16)

$$\lim_{s\to\infty} v^{\vec{n}_s}(\alpha) = \lim_{s\to\infty} \sum_{\substack{c\in B_k^{\vec{n}_s} \\ S(\alpha)\leq S(c)}} \frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))}, 1@s)}{N(\emptyset, 1@s)} prob(c) \tag{8.14}$$

$$= \lim_{s\to\infty} \sum_{\substack{c\in B_k^{\vec{n}_s}\cap F_s \\ S(\alpha)\leq S(c)}} \frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))}, 1@s)}{N(\emptyset, 1@s)} prob(c) \tag{8.15}$$

$$= \lim_{s\to\infty} \sum_{\substack{c\in B_k^{\vec{n}_s}\cap F_s \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} \tag{8.16}$$

$$= \sum_{\substack{c\in B_k^{\vec{q}_m} \\ S(\alpha)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}|} \tag{8.17}$$

$$= u^{\vec{q}_m}(\alpha) \ . \tag{8.18}$$

Where used for (8.17) that $\lim_{s\to\infty} \vec{n}_s = \vec{q}_m$.
The approximation for $\alpha \in SD_{<k}$ then follows from the fact that we can always write $w(\alpha)$ as the finite sum of probabilities of all $\beta \in SD_k$ extending α. □

Chapter 9

The $\eta-$Representation of a $u^{\vec{p}}$

It will follow from Lemma 28 on page 97 and Theorem 15 on page 96 that for $t \geq 2$ a $t-$heterogeneous probability function on a not purely unary language can never be language invariant. However there are $\leq t-$heterogeneous functions that are language invariant. For instance the $u^{\vec{p}}$ for $\vec{p} \in \mathbb{H}_t$ are such functions. Here we now calculate the $\eta-$representation [see theorem 5 on page 52] of the $u^{\vec{p}}$ which is a convex combination of heterogeneous functions.

If $p_0 > 0$ or if there are infinitely many colors in our urn then $u_L^{\vec{p}}$[1] is homogeneous hence the $\eta-$representation of such an $u^{\vec{p}}$ is just $u^{\vec{p}} = \eta_0 u^{\vec{p}}$ since $\eta_0 = 1$ for not purely unary L.
If on the other hand $\vec{p} \in \mathbb{H}_1$ then $u^{\vec{p}} = w^{[1]}$ and again we get a very simple representation. We then have $u^{\vec{p}} = \eta_1 w^{[1]} = w^{[1]}$.
If L is purely unary and there is a black ball or infinitely many colors in the urn then $u_L^{\vec{p}}$ is $2^{m_1(L)}-$heterogeneous. So in this case $\eta_{2^{m_1(L)}} = 1$.
For the remainder of this chapter we will keep a not purely unary language L fixed and will henceforth suppress it in the notation.

Remark 7. Let $E \in ER_t^l$. We recall that $SD_q^t := \{\alpha \in SD_q \mid |\overrightarrow{S(\alpha)}| = t\}$ and from the canonical representation theorem on page 38 that $|SD_l^l| \cdot |ER_t^l| = |SD_t^l|$ and hence

$$|ER_t^l| = \frac{|SD_t^l|}{|SD_l^l|} \quad . \tag{9.1}$$

For a given $\vec{p} \in \mathbb{H}_t$ we let for $1 \leq i \leq l$ $e_i' := \sum_{j \in E_i} p_j$. Next we order the elements of this tuple $\vec{e}'$ in non-increasing order and call the result $\vec{e}$. Then $\vec{e}$ is in $\mathbb{H}_l$ and so

[1] For L not purely unary!

$v^{\vec{e}}$ is well defined where $v^{\vec{e}}$ is the $l-$heterogeneous probability function defined in equation (7.2) on page 61.
Please note that it is possible that for different $E, F \in ER_t^l$ $\vec{e}$ and $\vec{f}$ are the same. For example for $\vec{p} = \langle 0.3, 0.25, 0.25, 0.2\rangle$ and $1, 2 \in E_1, 3, 4 \in E_2$ and $1, 3 \in F_1, 2, 4 \in F_2$ we find $e_1' = p_1 + p_2$, $f_1' = p_1 + p_3$ and so $\vec{e} = \vec{f} = \langle 0.55, 0.45\rangle$.
We write $B_q^{\vec{e}}$ to denote the set of draws of q balls with respect to $\vec{e}$. We also let C^E to be those draws from the urn given by $\vec{e}$ that contain all $|\vec{e}|$ colors in that urn.

We now get[2]

Theorem 13. *For $\vec{p} \in \mathbb{H}_t$ with $\eta_l = |SD_t^l(L)|/|SD_t(L)|$*

$$u_L^{\vec{p}} = \sum_{l=1}^{t} \frac{|SD_t^l(L)|}{|SD_t(L)|} \frac{1}{|ER_t^l|} \sum_{E \in ER_t^l} v_L^{\vec{e}} \ .$$

Proof. Recall that C was defined as the set of draws of at least t balls that contain all colors and that $\lim_{q\to\infty} \sum_{c \in B_q \cap C} prob(c) = 1$ from equation (7.3) on page 66.[3] Furthermore we defined on page 52 in equation (5.27)

$$\gamma_1 := \lim_{q\to\infty} \sum_{\vec{x} \in X_q^1} N(\emptyset, \vec{x}) \cdot w(\vec{x}) \ .$$

So when $w = u^{\vec{p}}$ we get that

$$\begin{aligned}
\gamma_1 &= \lim_{q\to\infty} N(\emptyset, \langle q\rangle) \cdot u^{\vec{p}}(\langle q\rangle) = |SD_1| \cdot \lim_{q\to\infty} u^{\vec{p}}(\langle q\rangle) \\
&= |SD_1| \lim_{q\to\infty} \sum_{c \in B_q \cap C} Contr(\langle q\rangle \mid c) \cdot prob(c) \\
&= \frac{|SD_1|}{|SD_t|} \lim_{q\to\infty} \sum_{c \in B_q \cap C} prob(c) = \frac{|SD_1|}{|SD_t|} = \frac{|SD_t^1|}{|SD_t|}
\end{aligned}$$

since every draw contributes to a state description with spectrum $\langle q\rangle$. For general $2 \leq$

[2] Just for the purpose of stating the theorem once in its full form we denote the dependencies on the not purely unary language L.

[3] Until further mention in the second part of the proof everything is with respect to the given $\vec{p} \in \mathbb{H}_t$. So B_q is with respect to $\vec{p}$ and so is C.

$l \leq t$ we have from equation (5.31) on page 54 for any w satisfying **(SX)** that

$$\gamma_l := \lim_{q\to\infty} \sum_{\vec{x}\in X_q^l} N(\emptyset,\vec{x}) \cdot w^{(l-1)}(\vec{x}) = \left(\prod_{i=1}^{l-1} \frac{1}{1-\gamma_i}\right) \lim_{q\to\infty} \sum_{\alpha\in SD_q^l} w(\alpha) \ .^4$$

So here we get using Lemma 20 on page 74

$$\begin{aligned} \gamma_l &= \left(\prod_{i=1}^{l-1} \frac{1}{1-\gamma_i}\right) \lim_{q\to\infty} \sum_{\alpha\in SD_q^l} \sum_{\substack{c\in B_q \\ S(\alpha)\leq S(c)}} Contr(\alpha|c)prob(c) \\ &= \left(\prod_{i=1}^{l-1} \frac{1}{1-\gamma_i}\right) \lim_{q\to\infty} \sum_{\alpha\in SD_q^l} \sum_{\substack{c\in B_q\cap C \\ S(\alpha)\leq S(c)}} Contr(\alpha|c)prob(c) \ . \qquad (9.2) \end{aligned}$$

Now we count how often a given draw $c \in B_q \cap C$ appears in (9.2). We therefore count the number of state descriptions in SD_q^l with an equivalence relation that is refined by $S(c)$. We will use the canonical representation which states that the number of state descriptions in SD_q with a fixed equivalence relation E equals $N(\emptyset, 1@|E|)$.
We need furthermore for a fixed $c \in B_q \cap C$ to calculate the number of equivalence relations $E \in ER_q^l$ that are refined by $S(c)$. This number however equals the number partitions of a set of size t into a set of size l. That is subsequently equal to $|SD_t^l|/|SD_l^l|$ as seen in (9.1). So for $c \in B_q \cap C$ c appears the following number of times in equation (9.2)

$$\begin{aligned} |\{\alpha \in SD_q^l \mid S(\alpha) \leq S(c)\}| &= |\{E \in ER_q^l \mid E \leq S(c)\}| \ \cdot \ |\{\alpha \in SD_q^l | S(\alpha) = E\}| \\ &= \frac{|SD_t^l|}{|SD_l^l|} \ \cdot \ N(\emptyset, 1@l) = \frac{|SD_t^l|}{|SD_l^l|} \ \cdot \ |SD_l^l| = |SD_t^l| \ . \end{aligned}$$

Whenever such a c appears in the sum above then the $Contr-$factor is $(|SD_{|S(c)|}|)^{-1}$

[4] As long as all the γ_i are strictly less than 1.

and so only depends on c and not on α. Hence we find for γ_l

$$\begin{aligned}\gamma_l &= \left(\prod_{i=1}^{l-1}\frac{1}{(1-\gamma_i)}\right)|SD_t^l|\lim_{q\to\infty}\sum_{c\in B_q\cap C}\frac{prob(c)}{|SD_{|S(c)|}|}\\ &= \left(\prod_{i=1}^{l-1}\frac{1}{(1-\gamma_i)}\right)|SD_t^l|\lim_{q\to\infty}\sum_{c\in B_q\cap C}\frac{prob(c)}{|SD_t|}\\ &= \frac{|SD_t^l|}{|SD_t|}\cdot\prod_{i=1}^{l-1}\frac{1}{1-\gamma_i}\quad .\end{aligned}$$

Hence by definition (see equation (5.36) on page 56) $\eta_l = \gamma_l\prod_{i=1}^{l-1}(1-\gamma_i) = \frac{|SD_t^l|}{|SD_t|}$. We can now check that things are indeed as they should be

$$\eta_1+\cdots+\eta_t = \gamma_1+\gamma_2\cdot(1-\gamma_1)+\cdots+\gamma_t\prod_{i=1}^{t-1}(1-\gamma_i) = \sum_{l=1}^{t}\frac{|SD_t^l|}{|SD_t|} = 1\quad .$$

In the second half of the proof we work out what the $l-$heterogeneous part of $u^{\vec{p}}$ is. Before proceeding with the second half however we give some explanations as to what is going on.

We think of $E\in ER_t^l$ as a way of dividing our t colors in the urn into l not-empty sets of colors. We then merge, or *glue together*, all colors in these sets. This gives us l new "super colors", with probabilities as given by $\vec{e}'$. Unfortunately $\vec{e}'$ is in general not in $\mathbb{H}_l$. We will hence have to reorder $\vec{e}'$ to $\vec{e}$ to obtain an element of $\mathbb{H}_l$. Then $v^{\vec{e}}$ is well defined.

As we will show the $l-$heterogeneous part of $u^{\vec{p}}$ is $\sum_{E\in ER_t^l} v^{\vec{e}}/|ER_t^l|$. This is a probability function since all the $v^{\vec{e}}$ are and we then take a convex combination. In order to show that this probability function does the trick we only have to verify the algebraic expression in equation (9.3). Once we checked that the proof is complete.

The $1-$heterogeneous part has to be $w^{[1]}$ since there's only one $1-$heterogeneous function which is of the right form for $l=1$ since $|E_t^1|=1$.

For general $2\le l\le t$ we recall equation (5.29) from page 54

$$w^{[l]}(\vec{x}) := \lim_{q\to\infty}\sum_{\vec{y}\in X_q^l}\gamma_l^{-1}N(\vec{x},\vec{y})w^{(l-1)}(\vec{y})\quad .$$

Here for $w = u^{\vec{p}}$ we find upon substituting the definition of $w^{(l-1)}$ for $\alpha \in SD^{\leq l}$ that

$$\begin{aligned} w^{[l]}(\alpha) &= \frac{1}{\gamma_l}\left(\prod_{i=1}^{l-1}\frac{1}{1-\gamma_i}\right)\lim_{q\to\infty}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}} u^{\vec{p}}(\beta) \\ &= \frac{1}{\eta_l}\lim_{q\to\infty}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}} u^{\vec{p}}(\beta) \\ &= \frac{1}{\eta_l}\lim_{q\to\infty}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}}\sum_{\substack{c\in B_q\cap C\\ S(\beta)\leq S(c)}}\frac{prob(c)}{|SD_t|} \\ &= \frac{1}{|SD_t^l|}\lim_{q\to\infty}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}}\sum_{\substack{c\in B_q\cap C\\ S(\beta)\leq S(c)}} prob(c) \ . \end{aligned}$$

For $\alpha \in SD^{\leq l}$ we have since the $v^{\vec{e}}$ are $l-$heterogeneous[5]

$$\begin{aligned} \frac{1}{|ER_t^l|}\sum_{E\in ER_t^l} v^{\vec{e}}(\alpha) &= \frac{1}{|ER_t^l|}\lim_{q\to\infty}\sum_{E\in ER_t^l}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}} v^{\vec{e}}(\beta) \\ &= \frac{1}{|ER_t^l|}\lim_{q\to\infty}\sum_{E\in ER_t^l}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}}\sum_{\substack{c^E\in B_q^{\vec{e}}\cap C^E\\ S(\beta)\leq S(c^E)}}\frac{prob(c^E)}{|SD_l^l|} \\ &= \frac{|SD_l^l|}{|SD_t^l|}\frac{1}{|SD_l^l|}\lim_{q\to\infty}\sum_{E\in ER_t^l}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}}\sum_{\substack{c^E\in B_q^{\vec{e}}\cap C^E\\ S(\beta)\leq S(c^E)}} prob(c^E) \\ &= \frac{1}{|SD_t^l|}\lim_{q\to\infty}\sum_{E\in ER_t^l}\sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}}\sum_{\substack{c^E\in B_q^{\vec{e}}\cap C^E\\ S(\beta)\leq S(c^E)}} prob(c^E) \ . \end{aligned}$$

We want to prove that the two probability functions $\sum_{E\in ER_t^l} v^{\vec{e}}/|ER_t^l|$ and $w^{[l]}$ are the same.

Let w_1, w_2 be two probability functions on L such that for all $\psi \in QFSL$ we have $w_1(\psi) \leq w_2(\psi)$. Then $1 - w_1(\psi) = w_1(\neg\psi) \leq w_2(\neg\psi) = 1 - w_2(\psi) \leq 1 - w_1(\psi)$ and so $w_1(\psi) = w_2(\psi)$. Hence $w_1 = w_2$.

By our standard Lemma 4 on page 18 it's hence enough that $w_1(\alpha) \leq w_2(\alpha)$ for all

[5] The superscript in the following on c^E only serves as a graphical reminder that the draw is from an urn where we glued colors together according to E.

$\alpha \in SD$. Hence it remains to show that

$$\lim_{q\to\infty} \sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}} \sum_{\substack{c\in B_q^{\vec{p}}\cap C\\ S(\beta)\leq S(c)}} prob(c) \leq \lim_{q\to\infty} \sum_{E\in ER_t^l} \sum_{\substack{\beta\in SD_q^l\\ \beta\vDash\alpha}} \sum_{\substack{c^E\in B_q^{\vec{e}}\cap C^E\\ S(\beta)\leq S(c^E)}} prob(c^E)\ .^6 \tag{9.3}$$

We need one last bit of new notation. Consider a $c \in B_q$ and a given $E \in ER_t^l$. c is draw of balls of at most t colors. We here think of E as a way of identifying some of these colors. We will replace a ball in c of color i with respect to $\vec{p}$ by a ball with color $vip_E(i)$ with respect to $\vec{e'}$. Informally $E(c)$ will be obtained from c by substituting every ball in that way.

Now for the precise definition let $E(c) \in B_q^{\vec{e}}$ be the unique draw in $B_q^{\vec{e}}$ that satisfies the following condition for all $1 \leq k \leq q$. Let i_k be the color of the $k-$th ball in c then the ball at the $k-th$ position of $E(c)$ has color n with respect to $\vec{e'}$ where n is such that $i_k \in E_n$ or equivalently $vip_E(i_k) = n$.

For example if $E_3 = \{4, 7\}$ then we replace all balls with color 7 or 4[7] in c by the ball with color 3 with respect to $\vec{e'}$. Note that $E(c)$ contains at most $l = |E|$ colors. If c contains t colors then $E(c)$ does contain l colors.

Now fix an $E \in ER_t^l$ and a $\beta \in SD_q^l$ extending α and consider a $c^E \in B_q^{\vec{e}} \cap C^E$ such that $S(\beta) = S(c^E)$. Then $prob(c^E) = e'_{i_1} \cdot \ldots \cdot e'_{i_q}$ for some $1 \leq i_1, \ldots, i_q \leq l$. We find since for all i $e'_i = \sum_{j\in E_i} p_j$ that

$$prob(c^E) = e'_{i_1} \cdot \ldots \cdot e'_{i_q} = \sum_{\substack{1\leq j_1,\ldots,j_q\leq t\\ j_1\in E_{i_1},\ldots,j_q\in E_{i_q}}} p_{j_1} \cdot \ldots \cdot p_{j_q} = \sum_{\substack{c\in B_q\\ S(\beta)\leq S(c)\\ E(c)=c^E}} prob(c)\ . \tag{9.4}$$

Now for a fixed $\beta \in SD_q^l$ extending α and draw $c \in B_q \cap C$ such that $S(\beta) \leq S(c)$ we consider the $E \in ER_t^l$ satisfying

$$S(\beta)ik \text{ implies } Ers$$

where c_r is the color of the $i-th$ ball in c and c_s is the color of the $k-th$ ball in c. That is E "glues" those colors together that lead to indistinguishable constants in β. Observe that for every such $c \in B_q \cap C$ there is precisely one such $E \in ER_t^l$ since

[6] We could of course show $\geq$ instead. However we feel that the proof for $\leq$ is less confusing.

[7] with respect to $\vec{p}$

$S(\beta) \in ER^l_q$. By (9.4) we find for such a $c \in B_q \cap C$ that

$$\sum_{\substack{d \in B_q \\ E(c)=E(d)}} prob(d) = prob(E(c)) \ \ .$$

Note that $S(E(c)) = S(\beta)$ since E is *the* "gluing" the appropriate colors together and $E(c)$ contains all l colors according to E since c contained all t colors with respect to $\vec{p}$. Hence we have

$$\lim_{q\to\infty} \sum_{\substack{\beta \in SD^l_q \\ \beta \models \alpha}} \sum_{\substack{c \in B_q \cap C \\ S(\beta) \leq S(c)}} prob(c) \leq \lim_{q\to\infty} \sum_{\substack{\beta \in SD^l_q \\ \beta \models \alpha}} \sum_{E \in ER^l_t} \sum_{\substack{c^E \in B^{\vec{e}}_q \cap C^E \\ S(\beta) \leq S(c^E)}} prob(c^E) \ \ .$$

□

Chapter 10

Including Equality

A good deal of the following is to appear in [18].

So far we were interested in the probability of spectra of state descriptions α. These spectra describe as to how the constants in α behave with respect to indistinguishability. The inquisitive reader might have asked himself why we didn't make any use of the notion of equality in first order logic. We now set out to explore this very idea.

10.1 Extending our Framework

We add the equality symbol $\equiv$ to our language with the usual convention for building well formed formulae. It is convenient to use the non-standard notation $\equiv^0\ a_i a_j$ for $a_i \not\equiv a_j$ (or the more pedantic $\neg\ a_i \equiv a_j$) and $\equiv^1\ a_i a_j$ for $a_i \equiv a_j$. We make the convention that $\equiv$ is not a relation symbol.[1]

Definition 46. Let L be a language in our sense. Then denote by $L^=$ the language obtained form L by adding $\equiv$ with the standard conventions of building well formed formulae. The set of sentences over $L^=$ will be denoted by $SL^=$ furthermore we define $QFSL^=$ to be the set of quantifier free sentences over $L^=$.
Let $QFS=$ be the fragment of $QFSL^=$ that doesn't contain any relation symbols. $QFS=$ obviously doesn't depend on the originally chosen language L. Denote by $S=$ the sentences of this language.

[1]We use the more pedantic $\equiv$ as symbol of first order logic to differentiate in the notation between a first order formulae and for instance statements about probabilities. Also we want to write things like $\varphi = a_1 \equiv a_2$ meaning that φ is the formula stating that a_1 and a_2 are the same. To reduce the amount of confusion the $\equiv$ −symbol comes in handy here.

From now on we we will consider the quantifier free **Axioms of Equality** to be tautologies. These axioms are

- A1) $a_i \equiv a_i$ for all i.
- A2) $a_i \equiv a_k \to a_k \equiv a_i$ for all i, k.
- A3) $(a_i \equiv a_k \wedge a_k \equiv a_s) \to a_i \equiv a_s$ for all i, k, s.
- A4) $(a_i \equiv a_k) \to (R\vec{a}a_k\vec{a}' \leftrightarrow R\vec{a}a_i\vec{a}')$ for all i, for all relation symbols and all tuples $\vec{a}, \vec{a}'$.

A1 to A4 will be referred to for short as the *axioms of equality*. From now on we mean by *consistent* always "consistent together with the axioms of equality".

Definition 47. A map $w : SL^{=} \to [0, 1]$ is called a *probability function on* $L^{=}$ if the following three conditions hold for all $\theta, \varphi, \exists x\psi(x) \in SL^{=}$:
(P1) If $\models \theta$ then $w(\theta) = 1$.
(P2) If $\models \neg(\theta \wedge \varphi)$ then $w(\theta \vee \varphi) = w(\theta) + w(\varphi)$.
(P3) $w(\exists x\psi(x)) = \lim_{m\to\infty} w(\bigvee_{i=1}^{m} \psi(a_i))$.

Mutatis mutandis extend the above definition to the language that only contains the equality symbol but no relation symbol.
With the above convention (P1) implies that each axiom of equality has to be assigned probability 1.

Lemma 23. For $\varphi, \psi \in SL^{=}$ and every probability function w on $SL^{=}$ it holds that

1. $w(\neg\varphi) = 1 - w(\varphi)$.
2. If $\models \varphi$ then $w(\neg\varphi) = 0$.
3. If $\models \varphi \to \psi$ then $w(\varphi) \leq w(\psi)$.
4. If $\models \varphi \leftrightarrow \psi$ then $w(\varphi) = w(\psi)$.
5. $w(\varphi \vee \psi) = w(\varphi) + w(\psi) - w(\varphi \wedge \psi)$.

Proof. The proofs for SL can be used here. Simply replace SL by $SL^{=}$ everywhere. See for instance [23] page 10. □

Gaifmans' theorem 1 as quoted on page 17 is actually stated in [8] for languages that may contain equality. So we have

Theorem 14. *Any function $v : QFSL^{=} \to [0,1]$ satisfying (P1) and (P2) for quantifier free sentences of $L^{=}$ and giving each instance of the axioms of equality value 1 extends uniquely to a probability function on $SL^{=}$.*

Hence we can concentrate from now on on $QFSL^{=}$, very much in the same fashion as we did when there was no equality symbol in the language.
Gaifmans theorem also holds in the special case where we consider the language containing only the equality symbol and no relation symbol.

Definition 48. An *equality state η on p over $L^{=}$* is a formula of the form

$$\eta = \alpha \wedge \Big(\bigwedge_{i,j=1}^{p} \equiv^{\epsilon_{ij}} a_i a_j \Big)$$

where $\alpha \in SD_p(L)$. The set of all consistent equality states on p will be denoted by $EQ_p(L)$. The union of all $EQ_p(L)$ is denoted by $EQ(L)$. We will drop the mention of L whenever possible.[2]
The $\bigwedge_{i,j=1}^{p} \equiv^{\epsilon_{ij}} a_i a_j$ part of an equality state on p will be called a *table on p*. The set of all consistent tables on p is denoted by T_p and the union over all p is denoted by T.
From now on for $\eta \in EQ_p$ with $\eta = \alpha \wedge \tau$ we always mean that $\tau \in T_p$ and $\alpha \in SD_p$.

Remark 8. Let $\tau \in T_p$ with $\tau = \bigwedge_{ik}^{p} \equiv^{\epsilon_{ij}} a_i a_j$. Since τ is consistent ϵ is a symmetric two-dimensional zero-one array on p with $\epsilon_{ii} = 1$. Furthermore for all $1 \leq i \leq p$ and for all $1 \leq i, j, k \leq p$ we have $\tau \vDash (a_i \equiv a_j \wedge a_j \equiv a_k) \to a_i \equiv a_k$.

Definition 49. For $\phi \in T_p$ or $\phi \in EQ_p$ we define an equivalence relation $S(\phi)$ on p. We set $S(\phi)ik$ to hold for $1 \leq i, k \leq p$ if and only if $\phi \vDash a_i \equiv a_k$.[3]
We let T_p^t be the subset of T_p containing all those tables with an equivalence relation with t classes and similarly EQ_p^t. Similarly $T_p^{<t}$, $EQ_p^{\geq t}$ and so on.

As we are going to see it makes sense to talk about $S(\tau)$ and $S(\eta)$ without invoking **(SX)**, giving us a strong hint that **(SX)** has some deep justification.

Lemma 24. Let $\eta \in EQ_p$ with $\eta = \alpha \wedge \tau$. Then $S(\alpha) \leq S(\tau)$.

Proof. This follows directly from the axioms of equality and the consistency of η. □

Corollary 6. Let $\tau \in T_p^t$. Then $|\{\alpha \in SD_p \mid \alpha \wedge \tau \text{ is consistent}\}| = |SD_t|$.

[2]Please note that an η_i in the $\eta-$representation of a probability function and an equality state η are completely different creatures.
[3]We want to mention that $\phi \vDash a_i \equiv a_k$ if and only if $\phi(a_i \mapsto a_k)$ is consistent.

Proof. $\alpha \wedge \tau$ is consistent if and only if $S(\alpha) \leq S(\tau)$. Looking up Lemma 11 on page 39 concludes the proof. □

Lemma 25. Let $f : EQ \to \{r \in \mathbb{R} \mid r \geq 0\}$ such that $\sum_{\eta \in EQ_1} f(\eta) = 1$ and such that for all p and all $\eta \in EQ_p$ we have

$$f(\eta) = \sum_{\substack{\eta' \in EQ_{p+1} \\ \eta' \vDash \eta}} f(\eta') \ .$$

Then f extends uniquely to a probability function on $QFSL^=$.

Proof. Use the same proof as in Lemma 4 on page 18, but replace in this proof SD everywhere by EQ and $QFSL$ by $QFSL^=$. The lemmata used to prove Lemma 4 hold here mutatis mutandis. □

Corollary 7. Let $f : T \to \{r \in \mathbb{R} \mid r \geq 0\}$ such that $\sum_{\tau \in T_1} f(\tau) = 1$ and such that for all p and all $\tau \in T_p$ we have

$$f(\tau) = \sum_{\substack{\rho \in T_{p+1} \\ \rho \vDash \tau}} f(\rho) \ .$$

Then f extends uniquely to a probability function on QFS= .

Proof. This a special case of the above lemma of a language without relation symbols. □

Remark 9. Let $\tau \in T_p$. Then τ is completely characterized by $S(\tau) \in ER_p$.
To recapture τ from $S(\tau)$ simply put $\epsilon_{ik} = 1$ if $S(\tau)ik$ holds and put it to 0 otherwise. Then $\tau = \bigwedge_{i,k=1}^{p} \equiv^{\epsilon_{ik}} a_i a_k$ as it should be. So for a fixed $E \in ER_p$ we have that $|\{\tau \in T_p \mid S(\tau) = E\}| = 1$.

Lemma 26. Let $\alpha \in SD_p^p(L)$ and let w be a probability function on $L^=$. Then there's only one $\tau \in T_p$ such that $\alpha \wedge \tau$ is consistent. Furthermore $w(\alpha) = w(\alpha \wedge \tau)$ for this table.

Proof. Since we can tell all constants over α apart the same has to hold for τ. But there's only one $\tau \in T_p$ with spectrum $1@p$. This τ satisfies $\tau = \bigwedge_{i,j=1}^{p} \equiv^{\delta_{ij}} a_i a_j$ where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.
From A4 we have that $\models \alpha \leftrightarrow (\alpha \wedge \tau)$. And since probability functions have to respect logical equivalences we have $w(\alpha) = w(\alpha \wedge \tau)$. □

Please note that this last observation holds in the absence of all the rationale principles we investigate. It is ”just” a consequence of the way we set up our functions assigning probability.

Lemma 27. Let ψ be an equality state or table on p. Then for $1 \leq i, k \leq p$

ψ is consistent if and only if $\psi(a_i \mapsto a_k, a_k \mapsto a_i)$ is consistent.[4]

Proof. The consistency of a first order formula does not depend on the name of the constants it contains. □

The following is already implicit in text above.

Definition 50. Let $\phi \in EQ \cup T$. Then the spectrum of ϕ is defined as the spectrum of $S(\phi)$.

Note that for $\eta \in EQ$ with $\eta = \alpha \wedge \tau$ the spectrum of η equals the spectrum of τ since $S(\eta) = S(\tau) \geq S(\alpha)$.

10.2 The Principle of Spectrum Exchangeability and Languages containing Equality

From now on we adopt the convention that all equality states and tables considered are consistent.

Principle of **Spectrum Exchangeability (SX=)**
Let $\eta \in EQ(L)$ and let w be a probability function on $L^=$. Then $w(\eta)$ only depends on the spectrum of η.

We feel that this principle pushes the ideas behind the formulation of **(SX)** a step further. We are now assigning probability according to whether we can distinguish constants using $\equiv$. This process is a much more clear cut than indistinguishability over state descriptions since every pair of different constants that look the same over an equality state will do so over every extension regardless of whether we extend to an equality state on more constants or to a language with more relation symbols.
As we are going to see very shortly there’s a very clear connection between the probability functions satisfying the above principle and the notion of language invariance.

[4] $\psi(a_i \mapsto a_k, a_k \mapsto a_i)$ is defined as the result of swapping a_i and a_k in ψ.

For $\vec{p} \in \mathbb{B}$ define probability functions on $QFSL^{=}$ respectively $QFS{=}$ by

$$u^{\vec{p}}_{L=}(\alpha \wedge \tau) := \sum_{\substack{c \in B_p \\ S(\tau)=S(c)}} \frac{prob(c)}{|SD_{|S(c)|}(L)|} \tag{10.1}$$

$$u^{\vec{p}}_{=}(\tau) := \sum_{\substack{c \in B_p \\ S(\tau)=S(c)}} prob(c) \ . \tag{10.2}$$

We will now check that these are indeed probability functions. Therefore observe

$$\sum_{\eta \in EQ_1} u^{\vec{p}}_{L=}(\eta) = \sum_{\alpha \in SD_1} u^{\vec{p}}_{L=}(a_1 \equiv a_1 \wedge \alpha) = \sum_{\alpha \in SD_1} u^{\vec{p}}_{L}(\alpha) = 1$$
$$\sum_{\tau \in T_1} u^{\vec{p}}_{=}(\tau) = u^{\vec{p}}(a_1 \equiv a_1) = \sum_{c \in B_1} prob(c) = 1 \ .$$

We find for $\tau \in T_p$ that

$$\begin{aligned}
\sum_{\substack{\tau' \in T_{p+1} \\ \tau' \vDash \tau}} u^{\vec{p}}_{=}(\tau') &= \sum_{\substack{E \in ER_{p+1} \\ S(\tau)=E\upharpoonright_p}} \sum_{\substack{c' \in B_{p+1} \\ S(c')=E}} prob(c') \\
&= \sum_{\substack{c \in B_p \\ S(c)=S(\tau)}} \left(prob(c) \sum_{d \in B_1} prob(d) \right) \\
&= \sum_{\substack{c \in B_p \\ S(c)=S(\tau)}} prob(c) \\
&= u^{\vec{p}}_{=}(\tau) \ .
\end{aligned}$$

Let $\alpha \in SD_p^{\leq t}$ and $E \in ER_{p+1}^{t+1}$ with $S(\alpha) \leq E\upharpoonright_p$ and $p+1 \in V(E)$ then

$$|\{\beta \in SD_{p+1} | S(\beta) \leq E,\ \beta \vDash \alpha\}| = |\{\gamma \in SD_{t+1} | \gamma \vDash \alpha/E\}| = \frac{|SD_{t+1}|}{|SD_t|}$$

what follows from the canonical representation results (see Lemma 11 on page 39), Remark 1 on page 16 and $\alpha/E = \alpha/(E\upharpoonright_p)$.
For $\eta \in EQ_p^t$ with $\eta = \alpha \wedge \tau$ and $\tau' \in T_{p+1}^t$ with $\tau' \vDash \tau$ there is only one $\alpha' \in SD_{p+1}$ such that $\alpha' \vDash \alpha$ and $\alpha' \wedge \tau'$ is consistent. Since if $S(\tau')\ i\ (p+1)$ then $S(\alpha')\ i\ (p+1)$ has to hold.

We now find for $\eta \in EQ_p^t$ with $\eta = \alpha \wedge \tau$ that

$$
\begin{aligned}
\sum_{\substack{\eta' \in EQ_{p+1} \\ \eta' \vDash \eta}} u^{\vec{p}}_{L=}(\eta') &= \sum_{\substack{\tau' \in T_{p+1} \\ \tau' \vDash \tau}} \sum_{\substack{\beta \in SD_{p+1} \\ S(\beta) \le S(\tau') \\ \beta \vDash \alpha}} u^{\vec{p}}_{L=}(\beta \wedge \tau') \\
&= \sum_{\substack{\tau' \in T^t_{p+1} \\ \tau' \vDash \tau}} \sum_{\substack{\beta \in SD_{p+1} \\ S(\beta) \le S(\tau') \\ \beta \vDash \alpha}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = S(\tau')}} \frac{prob(c')}{|SD_t|} \\
&+ \sum_{\substack{\tau' \in T^{t+1}_{p+1} \\ \tau' \vDash \tau}} \sum_{\substack{\beta \in SD_{p+1} \\ S(\beta) \le S(\tau') \\ \beta \vDash \alpha}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = S(\tau')}} \frac{prob(c')}{|SD_{t+1}|} \\
&= \sum_{\substack{E \in ER^t_{p+1} \\ S(\tau) = E\upharpoonright_p}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = E}} \frac{prob(c')}{|SD_t|} \\
&+ \sum_{\substack{E \in ER^{t+1}_{p+1} \\ S(\tau) = E\upharpoonright_p}} \sum_{\substack{\beta \in SD_{p+1} \\ S(\beta) \le E \\ \beta \vDash \alpha}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = E}} \frac{prob(c')}{|SD_{t+1}|} \\
&= \sum_{\substack{c \in B_p \\ S(c) = S(\eta)}} \frac{prob(c)}{|SD_t|} \sum_{\substack{d \in B_1 \\ d \in c}} prob(d) \\
&+ \sum_{\substack{E \in ER^{t+1}_{p+1} \\ S(\tau) = E\upharpoonright_p}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = E}} \frac{prob(c')}{|SD_{t+1}|} |\{\beta \in SD_{p+1} | S(\beta) \le E,\ \beta \vDash \alpha\}| \\
&= \sum_{\substack{c \in B_p \\ S(c) = S(\eta)}} \frac{prob(c)}{|SD_t|} \sum_{\substack{d \in B_1 \\ d \in c}} prob(d) \\
&+ \sum_{\substack{E \in ER^{t+1}_{p+1} \\ S(\tau) = E\upharpoonright_p}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = E}} \frac{prob(c')}{|SD_{t+1}|} \frac{|SD_{t+1}|}{|SD_t|} \\
&= \sum_{\substack{c \in B_p \\ S(c) = S(\eta)}} \frac{prob(c)}{|SD_t|} \sum_{\substack{d \in B_1 \\ d \in c}} prob(d) + \sum_{\substack{E \in ER^{t+1}_{p+1} \\ S(\tau) = E\upharpoonright_p}} \sum_{\substack{c' \in B_{p+1} \\ S(c') = E}} \frac{prob(c')}{|SD_t|} \\
&= \sum_{\substack{c \in B_p \\ S(c) = S(\eta)}} \frac{prob(c)}{|SD_t|} \sum_{\substack{d \in B_1 \\ d \in c}} prob(d) + \sum_{\substack{c \in B_p \\ S(c) = S(\eta)}} \frac{prob(c)}{|SD_t|} \sum_{\substack{d \in B_1 \\ d \notin c}} prob(d) \\
&= u^{\vec{p}}_{L=}(\eta) \quad .
\end{aligned}
$$

Hence $u^{\vec{p}}_{=}$ and $u^{\vec{p}}_{L=}$ are probability functions.
That they satisfy **(CX)**[5] follows from $S(c(i \mapsto k, k \mapsto i)) = S(c)(i \mapsto k, k \mapsto i)$ and the fact that η is consistent if and only if $\eta(i \mapsto k, k \mapsto i)$ is.
Furthermore the restriction of $u^{\vec{p}}_{L=}$ to QFS= equals $u^{\vec{p}}_{=}$.

The idea behind $u^{\vec{p}}_{=},\ u^{\vec{p}}_{L=}$ is that we don't want to glue constants together that might later get split. So whenever we draw a new color the corresponding constant is unlike all other constants. So if the $i-$th ball is black then a_i is different from all other constants. Hence every class with more than two elements is generated by a color. So if $S(\eta)$ or $S(\tau)$ have k classes with at least two elements and $\vec{p}$ has less than k colors[6] then the probability of η respectively τ is zero.
We will see in Lemma 29 on page 98 that $u^{\vec{p}}_{=}$ is in some sense the limit of the $u^{\vec{p}}_{L}$.

A theorem in [18] states

Theorem 15. *Let w be a probability function on L satisfying **(SX)**. Then w has an extension $w_{L=}$ to $L^{=}$ satisfying **(SX=)** if and only if there's a measure μ on $\mathbb{B}$ such that*

$$w = \int_{\mathbb{B}} u^{\vec{p}}_{L} d\mu(\vec{p}) \ . \tag{10.3}$$

If L is not purely unary then the extension $w_{L=}$ of w is unique and

$$w_{L=} = \int_{\mathbb{B}} u^{\vec{p}}_{L=} d\mu(\vec{p}) \tag{10.4}$$

for the same measure μ from above.

It can be shown that then

Corollary 8. A probability function w on $QFSL^{=}$ satisfies **(SX=)** if and only if there is a measure μ on $\mathbb{B}$ such that

$$w = \int_{\mathbb{B}} u^{\vec{p}}_{L=} d\mu(\vec{p}) \ . \tag{10.5}$$

A probability function w' on QFS= satisfies **(CX)** if and only if there is a measure μ

[5] Strictly speaking we haven't stated **(CX)** for $QFSL^{=}$ or QFS= . What we mean here is that the probability functions are invariant under renaming constants. To show that a probability function on $QFSL^{=}$ or QFS= satisfies **(CX)** it is enough to show that any transposition of constants in an equality state respectively table does not change the probability.

[6] Recall that we have the convention that black is not a color.

on $\mathbb{B}$ such that

$$w' = \int_{\mathbb{B}} u^{\vec{p}}_{=} d\mu(\vec{p}) \quad . \tag{10.6}$$

Furthermore for the same prior μ w and w' assign the same probabilities to sentences in $QFS{=}$.
And if w' is probability function on $QFS{=}$ satisfying **(CX)**, then for every language L there is a unique w on $QFSL^{=}$ satisfying **(SX=)** that agrees with w' on $QFS{=}$. This w is given by the same prior as w'.

This means that we have a one to one correspondence between the following three sets

- The set of language invariant families satisfying **(SX)**.
- The set of probability functions on $QFS{=}$ satisfying **(CX)**.
- The set of probability functions on $QFSL^{=}$ satisfying **(SX=)** for any fixed language L.

Now one might very well be inclined to think[7] that if we are only interested in assigning probabilities according to indistinguishability of constants then we can concentrate all our efforts on probability functions on $QFS=$ satisfying **(CX)**. As pleasing as the discovery of the one to one correspondences might be it does not solve all problems we are interested in. For instance, all the results for language invariance and the extension from $QFSL$ to $QFSL^{=}$ are phrased using the $u^{\vec{p}}$. But what about the $v^{\vec{p}}$?

10.3 Heterogeneity and (SX=)

Lemma 28. Let v be a $t-$heterogeneous probability function on $QFSL$ for $t > 1$ and L not purely unary. Then there is no extension w of v to $QFSL^{=}$ satisfying **(SX=)**.

Proof. Let $\delta \in SD^s_t$ with $s < t$, $\alpha \in SD^t_t$ and $\tau \in T^t_t$. Let us assume that the extension w exists. Then

$$\lim_{p\to\infty} \sum_{\substack{\beta\in SD^s_p\\ \beta\vDash\delta}} w(\beta\wedge\tau) \le \lim_{p\to\infty} \sum_{\substack{\beta\in SD^s_p\\ \beta\vDash\delta}} w(\beta) = \lim_{p\to\infty} \sum_{\substack{\beta\in SD^s_p\\ \beta\vDash\delta}} v(\beta) \le \lim_{p\to\infty} \sum_{\beta\in SD^s_p} v(\beta) = 0 \quad .$$

[7] and find some champagne bottles to open

Hence $\lim_{p\to\infty}\sum_{\substack{\beta\in SD_p^s\\ \beta\vDash\delta}} w(\beta\wedge\tau)=0$ for all $1\leq s<t$. This now implies

$$0<v(\alpha)=w(\alpha\wedge\tau)=w(\delta\wedge\tau)=\lim_{p\to\infty}\sum_{\substack{\beta\in SD_p^t\\ \beta\vDash\delta}} w(\beta\wedge\tau)\ . \tag{10.7}$$

All those β in (10.7) say that at least one constant from $a_{t+1},a_{t+2},\dots,a_p$ is different from the first t constants and τ says that all the first t constants are different.
Next we want to write $w(\beta\wedge\tau)$ as a sum of probabilities of equality states on p. So for $\tau'\in T_p$ with $\beta\wedge\tau'$ consistent all the first t constants are different via τ'. A constant a_l that is different over β from all the t constants in δ can't be equal to one of thirst t constants over τ'. Were this the case we'd have $S(\tau')il$ for some $1\leq i\leq t$ say, then by the consistency of $\beta\wedge\tau'$ we'd have $S(\beta)il$ and hence a_l couldn't be new over β.
Hence there is an equality state with spectrum of length at least $t+1$ with non-zero probability. So for an $\eta'\in EQ_{t+1}$ with spectrum $1@t+1$ we have $w(\eta')>0$.
But then (SX=) implies for every $\alpha\in SD_{t+1}$ and for $\tau'\in T_{t+1}^{t+1}$ with spectrum $1@t+1$ that $w(\alpha\wedge\tau')>0$. So in particular for $\gamma\in SD_{t+1}^{t+1}$ we find $w(\gamma\wedge\tau')>0$.
Hence $v(\gamma)=w(\gamma)\geq w(\gamma\wedge\tau')>0$. Contradicting the $t-$heterogeneity of v. □

Now suppose L is purely unary and $t=2^{m_1(L)}$. Consider a $\vec{p}$ with either infinitely many colors or a black ball. Then $u_L^{\vec{p}}$ is $t-$heterogeneous however we know that an extension of $u_L^{\vec{p}}$ to $QFSL^=$ exists, namely $u_{L=}^{\vec{p}}$.
In the proof above we consider a $\gamma\in SD_{t+1}^{t+1}$. If L is purely unary and $t=2^{m_1(L)}$ then there is no such γ. This explains why in general the above proof only works for not purely unary languages.
However the proof does work for purely unary languages L if $t<2^{m_1(L)}$.

10.4 Language Invariance and Equality

There is a neat explanation as to why the notions of *Language Invariance* and that of *Extendability to* $L^=$ are so closely connected.

Lemma 29. Let $\vec{p}\in\mathbb{B}$ and let $\tau\in T_p$ then

$$u_=^{\vec{p}}(\tau)=\lim_{|L|\to\infty}\sum_{\substack{\alpha\in SD_p(L)\\ S(\alpha)=S(\tau)}} u_L^{\vec{p}}(\alpha)$$

where the limit is taken over languages with an increasing number of relation symbols.

Proof. Over ever bigger languages two different colors in a draw c lead with ever greater probability to two distinguishable constants. That is $|SD_p^p(L)|/|SD_p(L)|$ converges to one when the number of relation symbols in L tends to infinity. Conversely $|SD_p^{<p}(L)|/|SD_p(L)|$ converges to zero.
Let $|L|$ denote the number of relation symbols in L. Then using the notion of a *direct limit* we find for all t

$$\lim_{|L|\to\infty} \frac{|SD_t^t(L)|}{|SD_{t+1}(L)|} \leq \lim_{|L|\to\infty} \frac{|SD_t(L)|}{|SD_{t+1}(L)|} = 0 \tag{10.8}$$

and by the canonical representation we have that for a given $E \in ER_p^t$

$$|SD_t^t(L)| = |\{\alpha \in SD_p^t \mid S(\alpha) = E\}| \ . \tag{10.9}$$

For $\tau \in T_p$ we find for any language L that

$$\begin{aligned} u_=^{\vec{p}}(\tau) &= \sum_{\substack{c\in B_p \\ S(c)=S(\tau)}} prob(c) \frac{|SD_{|S(c)|}(L)|}{|SD_{|S(c)|}(L)|} \\ &= \lim_{|L|\to\infty} |SD_{|S(\tau)|}^{|S(\tau)|}(L)| \sum_{\substack{c\in B_p \\ S(\tau)=S(c)}} \frac{prob(c)}{|SD_{|S(c)|}(L)|} && (10.10) \\ &= \lim_{|L|\to\infty} |SD_{|S(\tau)|}^{|S(\tau)|}(L)| \sum_{\substack{c\in B_p \\ S(\tau)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}(L)|} && (10.11) \\ &= \lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L) \\ S(\alpha)=S(\tau)}} \sum_{\substack{c\in B_p \\ S(\tau)\leq S(c)}} \frac{prob(c)}{|SD_{|S(c)|}(L)|} && (10.12) \\ &= \lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L) \\ S(\alpha)=S(\tau)}} u_L^{\vec{p}}(\alpha) \end{aligned}$$

where (10.10) follows from $S(c) = S(\tau)$ and we used (10.8) to obtain (10.11) and (10.9) for (10.12). □

Theorem 16. *Let w be a probability function on $QFS=$ satisfying **(CX)** and let $\tau \in T_p$ then*

$$w(\tau) = \lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L) \\ S(\alpha)=S(\tau)}} w_L(\alpha) \ .$$

Where w_L is the unique language invariant probability function on L induced by w.

Proof. By Lemma 29 the limit in $\lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L)\\ S(\alpha)=S(\tau)}} u_L^{\vec{p}}(\alpha)$ exists and is a probability function.

Let μ be the de Finetti prior for w. The theorem of dominated convergence (see Appendix, Corollary 15 on page 151) then implies

$$\begin{aligned} w(\tau) &= \int_{\mathbb{B}} u_{=}^{\vec{p}}(\tau) d\mu \\ &= \int_{\mathbb{B}} \lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L)\\ S(\alpha)=S(\tau)}} u_L^{\vec{p}}(\alpha) d\mu \\ &= \lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L)\\ S(\alpha)=S(\tau)}} \int_{\mathbb{B}} u_L^{\vec{p}}(\alpha) d\mu \\ &= \lim_{|L|\to\infty} \sum_{\substack{\alpha\in SD_p(L)\\ S(\alpha)=S(\tau)}} w_L(\alpha) \ . \end{aligned}$$

Here we use a slight generalization of the corollary 15 on page 151 in the appendix. We replace the limit with a *direct limit* to capture the less straight forward limit $\lim_{|L|\to\infty}$.

□

Chapter 11

The Paris Conjecture

It has long been proposed that the following holds for fixed not purely unary languages L. We will assume **(SX)** throughout this chapter.

11.1 The Conjecture

Conjecture 1. The Paris Conjecture
Let $\vec{x} \in X_p^t$ and $\vec{y} \in X_p^{\geq t}$ such that for all $1 \leq l \leq t$ we have $\sum_{i=1}^{l} x_i \geq \sum_{i=1}^{l} y_i$.[1]
Then for all w that satisfy **(SX)** on L we have

$$w(\vec{x}) \geq w(\vec{y}) \quad . \tag{11.1}$$

In [19] it was proved that

Theorem 17. *If w is a language invariant probability function then the conjecture does hold.*

So for example for $\vec{x} \in X_p^t$ with $x_s \geq x_{s+1} + 1$ then for every $\vec{q} \in \mathbb{B}$

$$u^{\vec{q}}(\langle x_1, \ldots, x_{s-1}, x_s + 1, x_{s+1}, \ldots, x_t \rangle^2) \geq u^{\vec{q}}(\langle x_1, \ldots, x_s, x_{s+1} + 1, x_{s+2}, \ldots, x_t \rangle) \quad .$$

The proof uses a generalized Muirhead inequality stated and proved in [27]. For $u^{\vec{q}}$ assuming that $u^{\vec{q}}(\vec{x}) > 0$ the proof shows that the inequality is sharp if $\vec{q} \neq \langle \frac{1}{t} @ t \rangle$ and $u^{\vec{q}} \neq w_\infty$. If $\vec{q} = \langle \frac{1}{t} @ t \rangle$ then $u^{\vec{q}}(\vec{x}) = u^{\vec{q}}(\vec{y})$.

[1] Since $\vec{x}$ and $\vec{y}$ are in X_p we then also have $\sum_{i=1}^{t} x_i \geq \sum_{i=1}^{r} y_i$ for all $t < r \leq |\vec{y}|$.
[2] If $x_s = x_{s-1}$ we have to reorder the first s entries to obtain an element of X_{p+1}.

If $\vec{x}, \vec{y} \in X_p$ are as in the conjecture then for every purely unary language L and every probability function w on L satisfying **(SX)** we have $w(\vec{x}) \geq w(\vec{y})$. However if $\vec{x}, \vec{y} \in X_p$ are not of this form then there are a purely unary language L and probability functions w', w'' on L satisfying **(SX)** such that $w'(\vec{x}) < w'(\vec{y})$ and $w''(\vec{x}) > w''(\vec{y})$; see [26].
Hence this condition is the *only* one that in general allows us to compare probabilities of different spectra on p.

Given a probability function w on L we now introduce a two place *conditional probability* function on $SL \times SL$ denoted by $w(\cdot|\cdot)$ such that

$$w(\varphi|\psi) = \frac{w(\varphi \wedge \psi)}{w(\psi)} \tag{11.2}$$

for $\varphi, \psi \in SL$ and $w(\psi) > 0$. We think of $w(\varphi|\psi)$ as the probability of φ in case that ψ holds.[3]

Remark 10. Now that we stated the conjecture and the theorem we give some explanations. We assume that L is not purely unary here.
Suppose $\alpha \in SD^t_{p-1}$ and consider a $\beta \in SD^t_p$ with spectrum $\vec{x}$ extending α where a_p is not a singleton over $S(\beta)$. Suppose for instance a_p is in the $i-$th class of $S(\beta)$ and this class has e elements.
Next consider a $\gamma \in SD^{\geq t}_p$ with spectrum $\vec{y}$ extending α such that a_p either joins a class that has less than e elements or forms a class on its own; in that case a_p might also split classes.
Then for all $1 \leq l \leq t$ we have $\sum_{k=1}^{l} x_k \geq \sum_{k=1}^{l} y_k$.
Assuming $w(\alpha) > 0$ and that the conjecture holds we get

$$w(\gamma|\alpha) = \frac{w(\gamma)}{w(\alpha)} \leq \frac{w(\beta)}{w(\alpha)} = w(\beta|\alpha)\ \ .$$

That means that if we assume that α is the state of the world then a_p is at least as likely to join a class with more elements. This allows us to make statements for the yet *unobserved* a_p. Such *inductive* conclusions have long been the holy grail in inductive logic [see for instance [12] and [29]].
It is very pleasing that at least in the language invariant case we can make such inductive assertions.

[3] Finally we investigate the motivating example given in the introduction.

As we are going to see next the condition in the conjecture is also the only one in the polyadic case.

11.2 A stronger result? No!

Lemma 30. Let $\vec{x}, \vec{y} \in X_p$ such that there are $i < j$ with $\sum_{k=1}^{i} x_k > \sum_{k=1}^{i} y_k$ and $\sum_{k=1}^{j} x_k < \sum_{k=1}^{j} y_k$. Then for every not purely unary language L there are homogeneous w, w' on L such that

$$w(\vec{x}) > w(\vec{y}) \text{ and } w'(\vec{x}) < w'(\vec{y}) \enspace .$$

Proof. Let $\alpha, \beta \in SD_p$ have spectrum $\vec{x}$ respectively $\vec{y}$.
Without loss of generality we assume that the first x_1 constants are equivalent over $S(\alpha)$ and so are $a_{x_1+1}, \ldots, a_{x_1+x_2}$ over $S(\alpha)$ and so on, for β we assume that the first y_1 constants are equivalent over $S(\beta)$ and so are $a_{y_1+1}, \ldots, a_{y_1+y_2}$ over $S(\beta)$ and so on. Put $I := \sum_{k=i+1}^{|\vec{y}|} y_k$ and observe that $Z := \sum_{k=i+1}^{|\vec{x}|} x_k < I$.
Consider the following $\vec{q}$ given by $q_0 := \epsilon > 0$ and $q_1, \ldots, q_i$ equal $\frac{1-\epsilon}{i}$ with ϵ as a parameter.
Every draw $c \in B_p^{\vec{q}}$ with $Contr(\beta|c) > 0$ contains at least I black balls since the y_k are not increasing and balls of the same color lead to equivalent constants. Hence there is a constant $\xi \in \mathbb{R}$ depending on the language and on $\vec{y}$ such that $u^{\vec{q}}(\beta) < \xi \cdot \epsilon^I$.
Now consider the following draw $d \in B_p$ that contributes to $u^{\vec{q}}(\alpha)$. First we draw the first color x_1 times, then the second color x_2 times and so on up to the $i-$th color x_i times. The last Z balls are all black. That is $d = c_1@x_1, c_2@x_2, \ldots, c_i@x_i, black@Z$. This draw contains strictly less than I black balls.
So $prob(d) \geq \epsilon^{I-1} \cdot (1-\epsilon)^{p-Z}/i^{p-Z}$. Hence $Contr(\alpha|d)prob(d) \geq \chi \cdot \epsilon^{I-1}$ for some constant $\chi \in \mathbb{R}$ which again depends on the language.
Shrinking ϵ far enough ensures that $u^{\vec{q}}(\alpha) > u^{\vec{q}}(\beta)$. Hence we eventually find our w.
The second part is very similar to the first part; we can recycle the proof for the first case since we never used that $i < j$. We here put $J := \sum_{k=j+1}^{|\vec{x}|} x_k$ and observe that $Y := \sum_{k=j+1}^{|\vec{y}|} y_k < J$.
Consider the following $\vec{q}$ given by $q_0 := \epsilon > 0$ and $q_1, \ldots, q_j$ equal $\frac{1-\epsilon}{j}$ with ϵ as a parameter.
Every draw $c \in B_p^{\vec{q}}$ with $Contr(\alpha|c) > 0$ contains at least J black balls since the x_k are not increasing and balls of the same color lead to equivalent constants. Hence there

is a constant $\xi' \in \mathbb{R}$ depending on the language and on $\vec{x}$ such that $u^{\vec{q}}(\alpha) < \xi' \cdot \epsilon^J$.
Now consider the following draw $d \in B_p$ that contributes to $u^{\vec{q}}(\beta)$. First we draw the first color y_1 times, then the second color y_2 times and so on up to the $j-$th color y_j times. The last Y balls are all black. This draw contains strictly less than J black balls. So $prob(d) \geq \epsilon^{J-1} \cdot (1-\epsilon)^{p-Y}/j^{p-Y}$. Hence $Contr(\beta|d)prob(d) \geq \chi' \cdot \epsilon^{J-1}$ for some constant $\chi' \in \mathbb{R}$ which again depends on the language.
Shrinking ϵ far enough ensures that $u^{\vec{q}}(\alpha) < u^{\vec{q}}(\beta)$. Hence we eventually find our w'. □

Analyzing the proof of Lemma 30 we get the following

Corollary 9. For different $\vec{x}, \vec{y} \in X$ and for every not purely unary L we can find a homogeneous probability function w on L such that $w(\vec{x}) \neq w(\vec{y})$.

Proof. First assume that $\vec{x} \in X_p$ and $\vec{y} \in X_q$ for different p, q. Then $w_\infty(\vec{x}) \neq w_\infty(\vec{y})$ as the completely independent solution treats atomic formulae as statistically independent.
Now if $\vec{x}, \vec{y}$ are both in X_p and different then there is an i with $\sum_{k=1}^{i} x_k > \sum_{k=1}^{i} y_k$ or $\sum_{k=1}^{i} x_k < \sum_{k=1}^{i} y_k$.
We can now use Lemma 30 to find a homogeneous w giving $\vec{x}$ and $\vec{y}$ different probabilities. □

Overall we have proved that in the language invariant case the name *only-rule* is in the polyadic case again well suited. We have

Theorem 18. *Let $\vec{x}, \vec{y} \in X_p$ and let L be not purely unary. Then the following are equivalent:*

- *For all language invariant w on L we have $w(\vec{x}) \geq w(\vec{y})$.*
- *For all $1 \leq i \leq |\vec{x}|$ we have $\sum_{k=1}^{i} x_k \geq \sum_{k=1}^{i} y_k$.*

11.3 Language Invariance and the Conjecture - Revisited

The following section dates back to an idea in [6] (page 226). The authors call it *Analogieschluss.* Translated to English that roughly means *conclusion by analogy.*
We saw that the conjecture makes a statement for a given $\alpha \in SD_{p-1}(L)$ about the

relative probability of a not yet observed constant a_p.
As the conjecture holds for language invariant w we can now speculate about the following assertion. Let $\alpha \in SD_p(L)$ and let w_L be a member of a language invariant family and suppose $L \subset L'$. Then $\gamma \in SD_p(L')$ with $\beta \in SD_p(L' \setminus L)$ such that $\gamma = \alpha \wedge \beta$ should be the more likely the more β *looks like* α.
Intuitively speaking this means that if we have taken a sample (have some data) α as to how p objects behave with respect to indistinguishability then a new sample (second set of data) β of the same objects is most likely to resemble the first sample.
This argumentation can be made precise. Indeed we have the following

Lemma 31. Let w be language invariant and let $\alpha \in SD_p(L)$. Suppose $\beta \in SD_p(L' \setminus L)$ for some $L' \supset L$ is such that $S(\beta) \leq S(\alpha)$. Then for all $\delta \in SD_p(L' \setminus L)$

$$w_{L'}(\alpha \wedge \beta) \geq w_{L'}(\alpha \wedge \delta)\ .[4] \tag{11.3}$$

Proof. If $S(\beta) \leq S(\alpha)$ then $S(\alpha \wedge \beta) = S(\alpha)$ since we can't tell any constant over β apart that wasn't already different over α. Note that then trivially the spectrum of α is the same as the spectrum of $\alpha \wedge \beta$.
So if $S(\delta)$ is also refined by $S(\alpha)$ then $w_{L'}(\alpha \wedge \beta) = w_{L'}(\alpha \wedge \delta)$.
If on the other hand $S(\delta)$ is not refinement of $S(\alpha)$ then there are two constants that can be told apart over γ but not over α. Hence $S(\alpha) < S(\alpha \wedge \delta)$.
Let $\vec{x}$ denote the spectrum of α and $\vec{y}$ that of $\alpha \wedge \delta$. Then consider the following function $f : |\vec{y}| \to |\vec{x}|$ with $f(i) := j$ if $S_i(\alpha \wedge \delta) \subset S_j(\alpha)$. Then by the definition of f we have that $\sum_{i,f(i)=j} y_i = x_j$ for all $1 \leq j \leq |\vec{x}|$. This shows that for $\vec{x}, \vec{y}$ that for all $1 \leq j \leq |\vec{x}|$ that $\sum_{k=1}^{j} x_k \geq \sum_{k=1}^{j} y_k$ since the entries in $\vec{x}, \vec{y}$ are ordered in non-increasing order.
Now since $w_L, w_{L'}$ are part of a language invariant family we find using Theorem 17 on page 101

$$w_{L'}(\alpha \wedge \beta) = w_{L'}(\vec{x}) \geq w_{L'}(\vec{y}) = w_{L'}(\alpha \wedge \delta)\ . \tag{11.4}$$

□

Definition 51. Let p_∞ be the following element of $\mathbb{B}$ given by $p_0 = 1$. The name is due to $u^{p_\infty} = w_\infty$.

Remark 11. The inequality in Lemma 31 is sharp if and only if the following two conditions are satisfied.

[4] If L is purely unary, then w might belong to more than one language invariant family, see theorem 12 on page 77. The here presented result holds for all language invariant $w_{L'}$ extending w_L.

- $S(\delta)$ is not refined by $S(\alpha)$.
- Let μ be the de Finetti prior of w_L. Then $0 < \int_{\mathbb{B}\setminus\{p_\infty\}} u_L^{\vec{p}}(\alpha)d\mu$.

Proof. If $S(\delta)$ is refined by $S(\alpha)$ then as seen above $w_{L'}(\alpha \wedge \beta) = w_{L'}(\alpha \wedge \delta)$.
If the second condition fails then either $w_L(\alpha) = 0$ or all the contribution to $w_L(\alpha)$ comes from the completely independent solution. That means that for

$$\mathbb{Q} := \{\vec{p} \in \mathbb{B} | (0 < p_0 < 1) \text{ or } (p_0 = 0 \ \& \ p_{|S(\alpha)|} > 0)\} \tag{11.5}$$

we have $\int_{\mathbb{Q}} 1 d\mu = 0$ since we either need at least $|S(\alpha)|$ colors or a black ball to find a draw contributing to α. So if $\vec{p} \in (\mathbb{B}\setminus\mathbb{Q})\setminus\{p_\infty\}$ then $u_L^{\vec{p}}(\alpha) = 0$. Hence for $\psi \in \{\beta, \delta\}$

$$\begin{aligned} w_{L'}(\alpha \wedge \psi) &= \int_{\mathbb{B}} u_{L'}^{\vec{p}}(\alpha \wedge \psi) d\mu \\ &= \int_{\mathbb{Q}} u_{L'}^{\vec{p}}(\alpha \wedge \psi) d\mu + \int_{(\mathbb{B}\setminus\mathbb{Q})\setminus\{p_\infty\}} u_{L'}^{\vec{p}}(\alpha \wedge \psi) d\mu + \int_{\{p_\infty\}} w_\infty(\alpha \wedge \psi) d\mu \\ &= \int_{\{p_\infty\}} w_\infty(\alpha \wedge \psi) d\mu = \int_{\{p_\infty\}} w_\infty(\alpha \wedge \beta) d\mu = \int_{\{p_\infty\}} w_\infty(\alpha \wedge \delta) d\mu \ . \end{aligned}$$

Now we'll prove the other implication, that is we show that the inequality is sharp if those two conditions are satisfied.
For every $\vec{q} \in \mathbb{Q}$ there is a $c \in B_p^{\vec{q}}$ with $Contr(\alpha|c) > 0$. And for every $c \in B_p^{\vec{q}}$ with $Contr(\alpha \wedge \delta|c) > 0$ we have $Contr(\alpha \wedge \delta|c) = Contr(\alpha \wedge \beta|c)$ since $S(\alpha) = S(\alpha \wedge \beta) < S(\alpha \wedge \delta)$.
Let a_g and a_f be two constants that can be told apart over δ but not over α. Then there is draw in $d \in B_p^{\vec{q}}$ such that the $g-$th and the $f-$th ball have color c_1 and such that $Contr(\alpha \wedge \beta|d) > 0$. However no such draw d can contribute to $\alpha \wedge \delta$ since a_f and a_g are different over $\alpha \wedge \delta$. Hence $u_{L'}^{\vec{q}}(\alpha \wedge \beta) > u_{L'}^{\vec{q}}(\alpha \wedge \delta)$.

As above for $\vec{q} \in (\mathbb{B} \setminus \mathbb{Q}) \setminus \{p_\infty\}$ we have $u^{\vec{q}}_{L'}(\alpha \wedge \beta) = u^{\vec{q}}_{L'}(\alpha \wedge \delta) = 0$. Hence

$$\begin{aligned}
w_{L'}(\alpha \wedge \beta) &= \int_{\mathbb{B}} u^{\vec{p}}_{L'}(\alpha \wedge \beta) d\mu \\
&= \int_{\mathbb{Q}} u^{\vec{p}}_{L'}(\alpha \wedge \beta) d\mu + \int_{(\mathbb{B}\setminus\mathbb{Q})\setminus\{p_\infty\}} u^{\vec{p}}_{L'}(\alpha \wedge \beta) d\mu + \int_{\{p_\infty\}} w_\infty(\alpha \wedge \beta) d\mu \\
&= \int_{\mathbb{Q}} u^{\vec{p}}_{L'}(\alpha \wedge \beta) d\mu + \int_{\{p_\infty\}} w_\infty(\alpha \wedge \beta) d\mu \\
&> \int_{\mathbb{Q}} u^{\vec{p}}_{L'}(\alpha \wedge \delta) d\mu + \int_{\{p_\infty\}} w_\infty(\alpha \wedge \delta) d\mu \\
&= \int_{\mathbb{B}} u^{\vec{p}}_{L'}(\alpha \wedge \delta) d\mu = w_{L'}(\alpha \wedge \delta) \quad .
\end{aligned}$$

□

We conclude that the "conclusion by analogy" is correct in case of a language invariant probability function.

11.4 The Conjecture and Heterogeneous Probabilities Functions

It is reasonable to try to use the proofs that show that the conjecture holds for language invariant probability functions also for heterogeneous probability functions since the representation theorems look very similar. However there is one important difference. The $Contr-$factor in the language invariant case does *not* depend on the state description α but only on $|S(\alpha)|$. The $Cont-$factor however does in general depend on the spectrum of α.[5]
On the upside we know that for $t-$heterogeneous w that $w(\vec{x}) = 0$ for all $\vec{x} \in X^{>t}$.
Next we'll prove a few results for some special cases. The general Paris Conjecture in the $t-$heterogeneous case is still very much a conjecture. We will need the following

Lemma 32. Let $a, b \in \mathbb{R}$ be different and positive. Then for all $n \geq 2$ we have that $a^n + b^n > a^{n-1}b + ab^{n-1}$.

Proof. W.l.o.g. we assume $a > b \geq 0$. Then since $a^{n-1} - b^{n-1} > 0$ we have

$$a^n + b^n - a^{n-1}b - ab^{n-1} = (a-b)(a^{n-1} - b^{n-1}) > 0 \quad .$$

[5] We want to recall that this factor is $\frac{N(\overrightarrow{\mathcal{S}(\alpha/S(c))}, 1@t)}{N(\emptyset, 1@t)}$.

□

Lemma 33. Let $\vec{p} \in \mathbb{H}_t$ then the Paris conjecture holds for different $\vec{x}, \vec{y} \in X_q^t$ and $v^{\vec{p}}$.[6] The inequality is sharp if and only if $\vec{p} \neq \langle \frac{1}{t} @ t \rangle$.

In essence this problem already appeared in the literature, see [9] pp. 44, under the name of Muirhead's Theorem.

Proof. If $t = 1$ then there is nothing to prove as X_q^1 contains only one spectrum (for a reminder on the set of spectra see example 4 on page 25).
We recall that C is set of draws of balls containing all colors with respect to $\vec{p}$.
Let $\vec{x}, \vec{y} \in X_q^t$ be different and such that for all $1 \leq i \leq t$ we have $\sum_{k=1}^{i} x_k \geq \sum_{k=1}^{i} y_k$ and let $\alpha, \beta \in SD_q^t$ have spectrum $\vec{x}$ respectively $\vec{y}$. By induction it is enough to prove the lemma for spectra $\vec{x}, \vec{y}$ such that there are $1 \leq s < l \leq t-1$ with $\langle x_1, \ldots, x_{s-1}, x_s, x_{s+1}, \ldots, x_{l-1}, x_l, x_{l+1}, \ldots x_t \rangle = \langle y_1, .., y_{s-1}, y_s{+}1, y_{s+1}, .., y_{l-1}, y_l{-}1, y_{l+1}, \ldots, y_t \rangle$ where $x_s > x_{s+1} + 1$. We have

$$v^{\vec{q}}(\vec{x}) = \sum_{\substack{c \in B_p^{\vec{q}} \\ S(\alpha)=S(c)}} prob(c) Cont(\alpha|c) = \sum_{\substack{c \in B_p^{\vec{q}} \cap C \\ S(\alpha)=S(c)}} \frac{prob(c)}{N(\emptyset, 1@t)} \tag{11.6}$$

$$v^{\vec{q}}(\vec{y}) = \sum_{\substack{c \in B_p^{\vec{q}} \\ S(\beta)=S(c)}} prob(c) Cont(\beta|c) = \sum_{\substack{c \in B_p^{\vec{q}} \cap C \\ S(\beta)=S(c)}} \frac{prob(c)}{N(\emptyset, 1@t)} \quad . \tag{11.7}$$

Hence we have to compare $\sum_{\sigma \in S^t} \prod_{i=1}^{t} p_{\sigma(i)}^{x_i}$ with $\sum_{\sigma \in S^t} \prod_{i=1}^{t} p_{\sigma(i)}^{y_i}$. If $\vec{p} = \frac{1}{t}@t$ then all entries of $\vec{p}$ are the same and hence the sums are the same. From now on hence assume that $\vec{p} \neq \langle \frac{1}{t}@t \rangle$.
For every $\sigma \in S^t$ there is a unique $\sigma' \in S^t$ such that $\sigma(i) = \sigma'(i)$ for all $i \notin \{s, l\}$ and such that $\sigma(s) = \sigma'(l)$ and $\sigma(l) = \sigma'(s)$.[7] So it is enough to show that for every $\sigma \in S^t$ with $p_{\sigma(s)} \neq p_{\sigma(l)}$ that

$$\prod_{i=1}^{t} p_{\sigma(i)}^{x_i} + \prod_{i=1}^{t} p_{\sigma'(i)}^{x_i} > \prod_{i=1}^{t} p_{\sigma(i)}^{y_i} + \prod_{i=1}^{t} p_{\sigma'(i)}^{y_i} \quad . \tag{11.8}$$

If $p_{\sigma(s)} = p_{\sigma(l)}$ then we have equality in (11.8). However as not all the colors have the same probability at least for some permutations $\sigma \in S^t$ $p_{\sigma(s)} \neq p_{\sigma(l)}$ has to hold.

[6]Please note that here both spectra have to be on q and have the same length t and $|\vec{p}| = t$.
[7]Observe that were we to start with $\sigma' \in S^t$ we would find σ as the corresponding permuation.

We start with left hand side of equation (11.8) and find using Lemma 32 to obtain the inequality:[8]

$$\prod_{i=1}^{t} p_{\sigma(i)}^{x_i} + \prod_{i=1}^{t} p_{\sigma'(i)}^{x_i} = \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{x_i}\big)[p_{\sigma(s)}^{x_s} p_{\sigma(l)}^{x_l} + p_{\sigma'(s)}^{x_s} p_{\sigma'(l)}^{x_l}]$$

$$= \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{x_i}\big)[p_{\sigma(s)}^{x_s} p_{\sigma(l)}^{x_l} + p_{\sigma(l)}^{x_s} p_{\sigma(s)}^{x_l}]$$

$$= \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{x_i}\big) p_{\sigma(s)}^{x_l} p_{\sigma(l)}^{x_l} [p_{\sigma(s)}^{x_s - x_l} + p_{\sigma(l)}^{x_s - x_l}]$$

$$> \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{x_i}\big) p_{\sigma(s)}^{x_l} p_{\sigma(l)}^{x_l} [p_{\sigma(s)}^{x_s - x_l - 1} p_{\sigma(l)} + p_{\sigma(l)}^{x_s - x_l - 1} p_{\sigma(s)}]$$

$$= \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{x_i}\big)[p_{\sigma(s)}^{x_s - 1} p_{\sigma(l)}^{x_l + 1} + p_{\sigma(l)}^{x_s - 1} p_{\sigma(s)}^{1 + x_l}]$$

$$= \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{y_i}\big)[p_{\sigma(s)}^{y_s} p_{\sigma(l)}^{y_l} + p_{\sigma(l)}^{y_s} p_{\sigma(s)}^{y_l}]$$

$$= \big(2 \prod_{\substack{i=1 \\ i \notin \{s,l\}}}^{t} p_{\sigma(i)}^{y_i}\big)[p_{\sigma(s)}^{y_s} p_{\sigma(l)}^{y_l} + p_{\sigma'(s)}^{y_s} p_{\sigma'(l)}^{y_l}]$$

$$= \prod_{i=1}^{t} p_{\sigma(i)}^{y_i} + \prod_{i=1}^{t} p_{\sigma'(i)}^{y_i} \quad .$$

□

There exists a dramatic shortcut for this proof. However as we are going to use *this* proof shortly there's no avoiding all those sums.

Now here is the shortcut. After equation (11.7) we could have used that the conjecture

[8] We assumed that $p_{\sigma(s)} \neq p_{\sigma(l)}$ and we have that $x_s - x_l > 1$.

holds for language invariant functions and have concluded for $\vec{p} \neq \frac{1}{t}@t$ that

$$v^{\vec{p}}(\vec{x}) = \frac{1}{N(\emptyset, 1@t)} \sum_{\sigma \in S^t} \prod_{i=1}^{t} p_{\sigma(i)}^{x_i} = \frac{|SD_t|}{N(\emptyset, 1@t)} \sum_{\sigma \in S^t} \prod_{i=1}^{t} \frac{p_{\sigma(i)}^{x_i}}{|SD_t|} = \frac{|SD_t|}{N(\emptyset, 1@t)} u^{\vec{p}}(\vec{x})$$
$$> \frac{|SD_t|}{N(\emptyset, 1@t)} u^{\vec{p}}(\vec{y}) = v^{\vec{p}}(\vec{y}) \ .$$

If $\vec{p} = \frac{1}{t}@t$ then $v^{\vec{p}}(\vec{x}) = v^{\vec{p}}(\vec{y})$ as seen in the proof above. □

Corollary 10. Let $\vec{p} \in \mathbb{H}_t$ then the Paris conjecture holds for different $\vec{x}, \vec{y} \in X_q^{t-1}$ and $v^{\vec{p}}$.[9] The inequality is sharp if and only if $\vec{p} \neq \langle \frac{1}{t}@t \rangle$.

Proof. There are two subcases. We will first consider the case that $q \geq t+1$.
Let $\alpha, \beta \in SD_q^{t-1}$ have spectrum $\vec{x}, \vec{y}$. Since $v^{\vec{p}}$ is $t-$heterogeneous all measure is eventually concentrated on state descriptions with a spectrum of length t.
Consider a $c \in B_q^{\vec{p}} \cap C$, so c contains all t colors, such that $S(\alpha) \leq S(c)$. Then by definition $Cont(\alpha|c) = \frac{N(\overrightarrow{S(\alpha/S(c))}, 1@t)}{N(\overrightarrow{S(\emptyset}, 1@t)}$.[10] If this factor is different from zero then $S(\alpha/S(c))$ has to have $t-1$ classes as $\alpha \in SD^{t-1}$ and at least one class contains more than one element as it's an equivalence relation on t since c contains t colors. So $\overrightarrow{S(\alpha/S(c))} = \langle 2, 1@t-2 \rangle$.
$N(\langle 2, 1@t-2 \rangle, 1@t) = 0$ since we need a new constant to split the class containing two elements, but then we'd have a spectrum on $t+1$.
So whenever $c \in C$ we know that $Cont(\alpha|c) = 0 = Cont(\beta|c)$. As α, β both allow us to tell $t-1$ constants apart we can now concentrate only on those draws containing $t-1$ different colors.
By definition we sum over all c with $S(\alpha) \leq S(c)$ to obtain $v^{\vec{p}}(\alpha)$. As seen above we here only need to sum over all c with $S(\alpha) = S(c)$. And so

$$v^{\vec{p}}(\vec{x}) = \sum_{\substack{c \in B_q^{\vec{p}} \\ S(\alpha) = S(c)}} prob(c) Cont(\alpha|c) = \sum_{\substack{c \in B_q^{\vec{p}} \\ S(\alpha) = S(c)}} \frac{N(\langle 1@t-1 \rangle, \langle 1@t \rangle) prob(c)}{N(\emptyset, 1@t)}$$
$$v^{\vec{p}}(\vec{y}) = \sum_{\substack{c \in B_q^{\vec{p}} \\ S(\beta) = S(c)}} prob(c) Cont(\beta|c) = \sum_{\substack{c \in B_q^{\vec{p}} \\ S(\beta) = S(c)}} \frac{N(\langle 1@t-1 \rangle, \langle 1@t \rangle) prob(c)}{N(\emptyset, 1@t)} \ .$$

[9] Please note that here both spectra have to be on q and have the same length $t-1$ and $|\vec{p}| = t$.
[10] We want to remind the reader that α/E for some equivalence relation E was introduced in definition 29 on page 33 as α divided by E.

So we here have to compare $\sum_{\sigma\in S^t}\prod_{i=1}^{t-1} p_{\sigma(i)}^{x_i}$ with $\sum_{\sigma\in S^t}\prod_{i=1}^{t-1} p_{\sigma(i)}^{y_i}$.[11] The only difference to the lemma above is that here the product is over $1 \leq i \leq t-1$ and not $1 \leq i \leq t$.
However in the proof above we never made any use of that, the only property we used was that the product contains at least two factors. As this here the case as well we can mutatis mutandis use the proof of that lemma.
If $q \in \{t, t-1\}$ then there is only one spectrum in X_q^{t-1}. It is $\langle 2, 1@t-2\rangle$ respectively $\langle 1@t-1\rangle$. Hence there are no different state description in X_q^{t-1} satisfying the only-rule relieving us hence from the burden of proof. □

To conclude this section on the only-rule and $t-$heterogeneous functions we want to state the following lemma to give a flavor of how we can use $w(1@t+1) = 0$ in the $t-$heterogeneous case. In the proof we will leave out a not very illuminating counting argument.

Lemma 34. Let $r_0(L) \geq 3$ and let w be $t-$heterogeneous. Then for all $2 \leq p \leq t$ and all $\vec{x} \in X_p^{<p}$

$$w(1@p) < w(\vec{x}) \ . \tag{11.9}$$

Proof. We'll do induction on l with $l := t-p$. To start off the induction we consider $t = p$. Due to $t-$heterogeneity we have $w(1@t) = \sum_{\vec{z}\in X_{t+1}} N(1@t, \vec{z})w(\vec{z}) = t \cdot w(2, 1@t-1)$.
For $\vec{y} \in X_p^{<p}$ we have that since $r_0(L) \geq 3$

$$N(\vec{y}, \langle 2, 1@p-1\rangle) > p \ . \tag{11.10}$$

This can be seen by going through lots of different cases and then for each case start counting extensions. Since the language is at least ternery there are at least $(2^{m_{r_0}})^{((p+1)^{r_0})}/(2^{m_{r_0}})^{(p^{r_0})} \geq (2)^{(p+1)^3-p^3}$ state descriptions in SD_{p+1} extending any given state description in $\mathcal{G} \in SD_p$. In the proof of Lemma 20 on page 120 we will see that we can always find an $\mathcal{H} \in SD_{p+1}$ extending $\mathcal{G}$ with spectrum $1@p+1$. Here we use the fact that the entries in $(H_{r_0})_{i,k,p+1@r_0-2}$ for $1 \leq i, k \leq p+1$ can be used to obtain any spectrum $\mathcal{H}$ could possibly have. All other entries in $\mathcal{H}$ can then be chosen freely under the condition that they don't allow us to tell constants apart that so far looked the same. We of course here set the particular spectrum we wish to obtain to $\langle 2, 1@p-1\rangle$.

[11] If $t = 2$ then X_q^{t-1} contains only one spectrum, which is $\langle q\rangle$, hence there's nothing to prove.

This can be used to show that we can find at least $p+1$ state descriptions with spectrum $\langle 2, 1@p-1\rangle$ extending our given state description. Hence by (11.10) for all $\vec{y} \in X_t^{<t}$

$$w(1@t) = t \cdot w(2, 1@t-1) < N(\vec{y}, \langle 2, 1@t-1\rangle) \cdot w(2, 1@t-1) \leq w(\vec{y}) \ . \quad (11.11)$$

For the induction step assume that $l = t - p > 0$. Observe that for $\vec{y} \in X_p^{<p}$

$$\begin{aligned} w(1@p) &= \sum_{\vec{z} \in X_{p+1}} N(1@p, \vec{z}) w(\vec{z}) \\ &= p \cdot w(2, 1@p-1) + N(\langle 1@p\rangle, \langle 1@p+1\rangle) w(1@p+1) \\ w(\vec{y}) &= \sum_{\vec{z} \in X_{p+1}} N(\vec{y}, \vec{z}) w(\vec{z}) \\ &= N(\vec{y}, \langle 2, 1@p-1\rangle) \cdot w(2, 1@p-1) + \sum_{\substack{\vec{z} \in X_{p+1} \\ \vec{z} \neq \langle 2, 1@t-1\rangle}} N(\vec{y}, \vec{z}) w(\vec{z}) \ . \end{aligned}$$

We'll now compare these two sums. Both contain the same number of terms of which there are $|SD_{p+1}|/|SD_p|$.
By the induction hypothesis $w(1@p+1)$ is the least probability in those sums. Hence it remains to show that $N(\vec{y}, \langle 2, 1@p-1\rangle) > p$ which is the case by equation (11.10). □

11.5 The Paris Conjecture and (SX=)

In this section we assume the all probability functions satisfy of **(SX)** respectively **(SX=)**.
The following is in [18]. It can also be seen to be an easy consequence of the fact that Paris conjecture holds for language invariant probability functions [see Theorem 17 on page 101] and the *neat explanation* [see Lemma 29 on page 98].
We recall that a table on p is a consistent formula of the form $\bigwedge_{i,k=1}^{p} a_i \equiv^{\epsilon_{ik}} a_k$. By $\widehat{x}$ we denote the spectrum of a table and by $\underline{x}$ the spectrum of an equality state.

Theorem 19. *Let $\widehat{x}, \widehat{y} \in X_p^t$ such that for all j we have $\sum_{i=1}^{j} x_i \geq \sum_{i=1}^{j} y_i$. Then $w(\widehat{x}) \geq w(\widehat{y})$.*

Remark 12. We have to insist that $\widehat{x}, \widehat{y}$ have the same length. We have seen for instance that $u_{=}^{p_\infty}(\widehat{1@p}) = 1$.
Furthermore if $S(\tau)$ has t classes with a least two elements and $p_t = 0$, then $u_{=}^{\vec{p}}(\tau) = 0$.

That follows from the fact that a black ball always leads to a singleton in $S(\tau)$ and two different colored balls always lead two to constants that are different over τ.

Corollary 11. Let $\underline{x}, \underline{y} \in X_p^t$ such that for all j $\sum_{i=1}^{j} x_i \geq \sum_{i=1}^{j} y_i$. Then $w(\underline{x}) \geq w(\underline{y})$.

Proof. We have for $\tau \in T_p$ that

$$w(\tau) = \sum_{\substack{\eta \in EQ_p \\ \eta \models \tau}} w(\eta) \quad .$$

All those η have the same spectrum as τ and by **(SX=)** the same probability.
With the theorem above and Corollary 6 on page 91 we now find

$$w(\underline{x}) = \frac{w(\widehat{x})}{|SD_t|} \geq \frac{w(\widehat{y})}{|SD_t|} = w(\underline{y}) \quad .$$

□

Chapter 12

The Principle of Instantial Relevance

The following has long been considered *the* appropriate notion of inductive reasoning (it has has been suggested in [3], pp. 974-976).

The **Principle of Instantial Relevance**
Let $\alpha \in SD_q$, $\beta \in SD_{q+1}$ and let $\gamma \in SD_{q+2}$ such that $\beta \vDash \alpha$, $\gamma \vDash \beta$ and such that $S(\gamma)(q+1)\ (q+2)$. Then

$$w(\beta|\alpha) = \frac{w(\beta)}{w(\alpha)} \leq \frac{w(\gamma)}{w(\beta)} = w(\gamma|\beta)\ . \tag{12.1}$$[1]

This principle has long been under investigation for purely unary languages (see [11], [2] and [23] for a modern treatment). We will here now see what happens in the polyadic case. For the rest of this chapter L is always not purely unary.
The motivation to consider such a principle is again that we want to be able to make *inductive inferences.*

Lemma 35. Let w be $t-$heterogeneous and α, β, γ as above such that $|S(\alpha)| = |S(\beta)| = |S(\gamma)| = t$. Then

$$\frac{w(\beta)}{w(\alpha)} \leq \frac{w(\gamma)}{w(\beta)}\ .$$

The inequality is sharp if and only if $w \neq v^{\langle \frac{1}{t}@t \rangle}$.

Proof. It is enough to prove the sharp inequality for an arbitrary $v^{\vec{p}}$ and see that in case $\vec{p} = \langle \frac{1}{t}@t \rangle$ equality holds.
Let a_{q+1} be equivalent to a_k over γ with $k \leq q$. Let l be such that $a_k \in S_l(\alpha)$ and let $\vec{x}$

[1]Assuming we don't divide by zero. We will here always assume this.

be the spectrum of α. We now find

$$\begin{aligned}\frac{w(\gamma)\cdot w(\alpha)}{w(\beta)^2} &= \frac{(\sum_{\sigma\in S^t} p^2_{\sigma(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)})\cdot(\sum_{\sigma\in S^t}\prod_{i=1}^t p^{x_i}_{\sigma(i)})}{(\sum_{\sigma\in S^t} p_{\sigma(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)})\cdot(\sum_{\sigma\in S^t} p_{\sigma(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)})}\\ &= \frac{\sum_{\sigma\in S^t}\sum_{\sigma'\in S^t} p^2_{\sigma(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)}\cdot\prod_{i=1}^t p^{x_i}_{\sigma'(i)}}{\sum_{\sigma\in S^t}\sum_{\sigma'\in S^t} p_{\sigma(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)}\cdot p_{\sigma'(l)}\prod_{i=1}^t p^{x_i}_{\sigma'(i)}}\\ &= \frac{\sum_{\sigma\in S^t,\sigma'\in S^t} p^2_{\sigma(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)}\cdot\prod_{i=1}^t p^{x_i}_{\sigma'(i)}}{\sum_{\sigma\in S^t,\sigma'\in S^t} p_{\sigma(l)}p_{\sigma'(l)}\prod_{i=1}^t p^{x_i}_{\sigma(i)}\cdot\prod_{i=1}^t p^{x_i}_{\sigma'(i)}}\ .\end{aligned}$$

If $\vec{p} = \langle \frac{1}{t}@t\rangle$ then the expression above has value one and hence $w(\beta)/w(\alpha) = w(\gamma)/w(\beta)$.
We put $A_\sigma := \prod_{i=1}^t p^{x_i}_{\sigma(i)}$. It hence remains to show for $\vec{p} \neq \langle \frac{1}{t}@t\rangle$ that

$$\frac{\sum_{\sigma\in S^t,\sigma'\in S^t} p^2_{\sigma(l)}A_\sigma A_{\sigma'}}{\sum_{\sigma\in S^t,\sigma'\in S^t} p_{\sigma(l)}p_{\sigma'(l)}A_\sigma A_{\sigma'}} > 1\ .$$

We now fix different $\sigma, \sigma' \in S^t$ and find

$$\begin{aligned}\frac{p^2_{\sigma(l)}A_\sigma A_{\sigma'} + p^2_{\sigma'(l)}A_\sigma A_{\sigma'}}{p_{\sigma(l)}p_{\sigma'(l)}A_\sigma A_{\sigma'} + p_{\sigma'(l)}p_{\sigma(l)}A_\sigma A_{\sigma'}} &= \frac{p^2_{\sigma(l)} + p^2_{\sigma'(l)}}{2p_{\sigma(l)}p_{\sigma'(l)}}\\ &\begin{cases} = 1 & \text{if } p_{\sigma(l)} = p_{\sigma'(l)},\\ > 1 & \text{if } p_{\sigma(l)} \neq p_{\sigma'(l)}\ .\end{cases}\end{aligned}$$

If $\vec{p} \neq \frac{1}{t}@t$ then at least for one pair of different $\sigma, \sigma' \in S^t$ we have that $p_{\sigma(l)} \neq p_{\sigma'(l)}$. Hence $v^{\vec{p}}(\beta)/v^{\vec{p}}(\alpha) < v^{\vec{p}}(\gamma)/v^{\vec{p}}(\beta)$. □

Corollary 12. Under the same conditions for α, β, γ we have for $u^{\vec{p}}$ with $\vec{p} \in \mathbb{H}_t$ and $|S(\alpha)| = t$ that

$$\frac{u^{\vec{p}}(\beta)}{u^{\vec{p}}(\alpha)} \leq \frac{u^{\vec{p}}(\gamma)}{u^{\vec{p}}(\beta)}\ .$$

Again equality holds if and only if $\vec{p} = \langle \frac{1}{t}@t\rangle$.

Proof. This follows directly from the lemma above since for any $\vec{x} \in X^t$ we have $|SD_t| \cdot u^{\vec{p}}(\vec{x}) = N(\emptyset, 1@t) \cdot v^{\vec{p}}(\vec{x})$. □

Corollary 13. Let $\beta, \gamma, t, \vec{p}, L$ be as in the above lemma. Let $\alpha \in SD^{t-1}_q$ such that $\beta \vDash \alpha$. Then

$$\frac{v^{\vec{p}}(\beta)}{v^{\vec{p}}(\alpha)} < \frac{v^{\vec{p}}(\gamma)}{v^{\vec{p}}(\beta)}\ . \qquad (12.2)$$

Proof. We recall that $Inj(n,t)$ denotes the set of injective maps from $\{1,\ldots,n\}$ into $\{1,\ldots,t\}$. There's an obvious one to one correspondence between $Inj(t-1,t)$ and $Inj(t,t)$ since every injective map $f \in Inj(t-1,t)$ can be extended uniquely to a $g \in Inj(t,t)$ by setting $g(i) := f(i)$ for all $1 \leq i \leq t-1$ and $g(t)$ the unique element of $\{1,\ldots,t\}$ not in the image of f. Hence $|Inj(t-1,t)| = |Inj(t,t)|$.
Let $\vec{z}$ be the spectrum of γ, n be such that $z_n = 2$ and let $\vec{x}$ be the spectrum of α. We note that for the spectrum $\vec{y}$ of β we have $\vec{y} = \langle x_1,\ldots,x_{t-1},1\rangle$ and $\vec{z} = \langle x_1,\ldots,x_{n-1},2,x_n,\ldots,x_{t-1}\rangle$. We have seen above that only draws with $t-1$ different colors contribute to the probability of α (see Corollary 10 on page 110). Their contribution is $N(1@t-1, 1@t)/N(\emptyset, 1@t)$. With $N(t) := N(\emptyset, 1@t)$ we now find

$$\begin{aligned}
&\frac{w(\gamma)\cdot w(\alpha)}{w(\beta)^2} \\
&= \frac{(\sum_{f\in Inj(t,t)} p_{f(t)}^2 \prod_{i=1}^{t-1} p_{f(i)}^{x_i})\cdot(\sum_{h\in Inj(t-1,t)} \prod_{i=1}^{t-1} p_{h(i)}^{x_i})}{(\sum_{g\in Inj(t,t)} p_{g(t)} \prod_{i=1}^{t-1} p_{g(i)}^{x_i})\cdot(\sum_{g\in Inj(t,t)} p_{g(t)} \prod_{i=1}^{t-1} p_{g(i)}^{x_i})} \frac{N(1@t-1,1@t)N(t)^2}{N(t)\cdot N(t)} \\
&= \frac{(\sum_{f\in Inj(t,t)} p_{f(t)}^2 \prod_{i=1}^{t-1} p_{f(i)}^{x_i})\cdot(\sum_{h\in Inj(t,t)} \prod_{i=1}^{t-1} p_{h(i)}^{x_i})}{(\sum_{g\in Inj(t,t)} p_{g(t)} \prod_{i=1}^{t-1} p_{g(i)}^{x_i})\cdot(\sum_{g'\in Inj(t,t)} p_{g'(t)} \prod_{i=1}^{t-1} p_{g'(i)}^{x_i})} N(1@t-1,1@t) \ .
\end{aligned}$$

In the proof of the lemma above we considered a term very similar to the first factor. There we considered products ranging from 1 to t. Here we have products from 1 to $t-1$. However in that proof we never made use of the fact that the products are from 1 to t. Hence we can conclude here directly that the first factor is greater or equal than one.
For the second factor observe that $N(\langle 1@t-1\rangle, \langle 1@t\rangle) = \frac{|SD_t|}{|SD_{t-1}|} - (t-1)$, here the new constant a_t either joins one of the $t-1$ classes or forms a class of its own. Hence $\frac{|SD_t|}{|SD_{t-1}|} - (t-1) > 1$. □

Furthermore we want to mention that the principle of instantial relevance does not hold for w_∞. Since

$$\frac{w_\infty(\beta)}{w_\infty(\alpha)} = \frac{|SD_p|}{|SD_{p+1}|} > \frac{|SD_{p+1}|}{|SD_{p+2}|} = \frac{w_\infty(\gamma)}{w_\infty(\beta)} \ . \tag{12.3}$$

If q_0 is close to 1 then $u_L^{\vec{q}}$ also fails to satisfy **(PIR)**. Since then for some $\alpha,\beta,\gamma \in SD$ as above the probability of $u^{\vec{p}}$ of these state descriptions is close enough to their probability under w_∞ that $u^{\vec{q}}(\beta)/u^{\vec{q}}(\alpha) > u^{\vec{q}}(\gamma)/u^{\vec{q}}(\beta)$.
We conjecture that implementing a correction term in **(PIR)** could overcome the failure

demonstrated above. We propose to consider

$$\frac{w(\gamma) \cdot w(\alpha)}{w(\beta)^2} \geq \frac{|SD_{p+1}(L)|^2}{|SD_{p+2}(L)| \cdot |SD_p(L)|} \quad . \tag{12.4}$$

This captures the fact that the number of literals to be decided to go from α to β is the same as going from β to γ for a purely unary language. Hence we should in case of a polyadic language correct by these factors.
For a purely unary language the here proposed version of **(PIR)** coincides with previous one as the correction terms vanish.

Remark 13. On a purely unary language **(CX)** implies **(PIR)** (see for instance [24]). For a polyadic language there is initially some hope to prove that $t-$heterogeneous functions satisfy **(PIR)**.
However we can now see that is not so.
Recall that we can approximate w_∞ on SD_{q+2} by sequence $(v^{\vec{p}_n})_{n\in\mathbb{N}}$ (see Lemma 22 on page 78). Since $\frac{w_\infty(\beta)}{w_\infty(\alpha)} > \frac{w_\infty(\gamma)}{w_\infty(\beta)}$ there is some huge n such that $\frac{v^{\vec{p}_n}(\beta)}{v^{\vec{p}_n}(\alpha)} > \frac{v^{\vec{p}_n}(\gamma)}{v^{\vec{p}_n}(\beta)}$.

Chapter 13

Johnson's Sufficientness Principle

13.1 The Principle

In this chapter we revisit an old principle that has been investigated intensively for unary languages. Results from these investigations were used to justify Carnap's famous continuum of probability functions c_λ with $\lambda \in [0, \infty]$. The c_λ are the only probability functions satisfying **(CX)** and Johnson's Sufficientness Principle[1] (see [23]). It has been proved in [28] that for a purely binary language only w_∞ and $w^{[1]}$ satisfy both these principles. We will here show the same for languages containing at least one relation symbol that is not unary nor binary. An analysis of the proof in [28] shows that this proof can be easily adapted to the case of a language containing binary and unary relation symbols.

Overall the following picture emerges: If L is purely unary and contains at least two relation symbols then the c_λ are the only functions satisfying both principles. If L is not purely unary then only w_∞ and $w^{[1]}$ do.

Since we surely want **(CX)** to hold we feel that based on the here presented results we can't accept Johnson's Sufficientness Principle in its here presented form.

Please note that all the here presented results hold in the absence of **(SX)**. We will however use some terminology that has been developed to make use of **(SX)**.

The following principle was first proposed by W.E. Johnson and now carries his name. It was later rediscovered by Carnap.

Johnson's Sufficientness Principle (JSP)

[1] If the purely unary language contains at least two relation symbols.

Let $\alpha \in SD_p$ with $w(\alpha \upharpoonright_{p-1}) > 0$ then $w(\alpha | \alpha \upharpoonright_{p-1})$ only depends on p and on how many constants are equivalent to a_p over α.

Definition 52. Let $\alpha \in SD_q$ and assume that w satisfies **(JSP)**. If $w(\alpha \upharpoonright_{q-1}) > 0$ then the following is well defined for all $2 \leq p \leq q$

$$g(p-1, s_p) := \frac{w(\alpha \upharpoonright_p)}{w(\alpha \upharpoonright_{p-1})} = w(\alpha \upharpoonright_p \ | \ \alpha \upharpoonright_{p-1}) \tag{13.1}$$

where $s_p := |\{1 \leq l \leq p-1 \mid S(\alpha \upharpoonright_p) lp\}|$.
For $p = 1$ we condition on the empty formula and have $g(0,0) = \frac{w(\alpha \upharpoonright_1)}{1}$. Since for all $\gamma \in SD_1$ $w(\gamma) = g(0,0)$ we find $g(0,0) = \frac{1}{|SD_1|} = w(\gamma)$.

Remark 14. Let $\alpha \in SD_p$ with $w(\alpha \upharpoonright_{p-1}) > 0$. By induction we now find $w(\alpha) = w(\alpha \upharpoonright_1) \cdot \prod_{i=2}^{p} w(\alpha \upharpoonright_i \ | \ \alpha \upharpoonright_{i-1})$. Hence

$$w(\alpha) = \prod_{i=0}^{p-1} g(i, s_i) \ . \tag{13.2}$$

The following was proved in [28]. An analysis of that proof shows that the lemma also holds if L contains binary and unary relation symbols.

Lemma 36. Let L be purely binary. Then for all $\alpha \in SD_p^p$ there exists a permutation $\pi \in S^p$ [2] such that for every $1 \leq k \leq p$ we have $\overrightarrow{\mathcal{S}(\alpha(\pi)) \upharpoonright_k} = \langle 1@k \rangle$.[3]

But if L contains only one relation symbol of a higher arity the lemma above fails! This is one of the apparently very rare differences between binary and polyadic languages.

Lemma 37. Let $r_0(L) \geq 3$. Then there exists an $\alpha \in SD_{r_0}^{r_0}$ such that for all permutation $\pi \in S^p$ and every $1 \leq k \leq r_0 - 1$ we have $\alpha(\pi) \upharpoonright_k \in SD_k^1$.

Proof. Let R be a relation symbol of arity r_0. Then put

$$\alpha := \neg R a_1 \ldots a_{r_0} \wedge \beta \tag{13.3}$$

where β is the conjunction of literals that makes α into a state description and β decides every literal it contains positively.
Since $\alpha \models \neg R a_1 \ldots a_{r_0} \wedge R a_1 \ldots a_{i-1} a_k a_{i+1} \ldots a_{r_0}$ for all $i \neq k$ the spectrum of α is

[2] depending on α
[3] $\alpha(\pi)$ was defined on page 45, definition 34.

$1@r_0$.
Whenever we delete all literals in α that contain a fixed constant we always delete $\neg Ra_1 \ldots a_{r_0}$. We are then left with a conjunction of literals ψ that is implied by β. Hence ψ does not contain a single negation symbol. So ψ is modulo renaming constants a state description with spectrum of length one. □

13.2 (JSP) and (CX) on a Polyadic Language

Definition 53. Let $\alpha \in SD_p$ and let R_1 be a relation symbols of maximal arity in L. We say that α is the identity and write $\alpha = ID_p$ if and only if α is of the form

$$\alpha = \bigwedge_{i=1}^{p} R_1 a_i \ldots a_i \wedge \nu$$

where ν is the conjunction of literals making α a state description that decides every literal negative. That is every every occurrence of every relation symbol in ν is negated. Note that the spectrum of ID_p for not purely unary languages is $1@p$, since for $1 \leq i < k \leq p$ we have $\alpha \models (R_1 a_i \ldots a_i \wedge \neg R_1 a_k a_i \ldots a_i)$.
The name derived from the fact that if L contains only one binary relation symbol then the two-dimensional array on p representing ID_p is the identity matrix.

Theorem 20. *Let $r_0(L) \geq 3$. Then the only probability functions satisfying* **(JSP)** *and* **(CX)** *are $w^{[1]}$ and w_∞.*

Proof. We assume that w satisfies **(JSP)** and **(CX)**. To kick things off recall that for all $\beta \in SD_1$ we have $w(\beta) = g(0,0) = \frac{1}{|SD_1|} > 0$.
Assume $g(1,0) = 0$ and assume that there exists a p such that $g(p,0) > 0$. We will show that this leads to a contradiction.
Then by **(JSP)** there exists an $\alpha \in SD_{p+1}^{\geq 2}$ such that $w(\alpha) > 0$ and such that a_{p+1} is a singleton over $S(\alpha)$ such that $S(\alpha \upharpoonright_{\{p,p+1\}} (a_p \mapsto a_1, a_{p+1} \mapsto a_2))$ consists of the two singletons $\{a_1\}$ and $\{a_2\}$. So we pick α such that a_{p+1} is new and such that we can distinguish a_p from a_{p+1} over $\alpha \upharpoonright_{\{p,p+1\}}$.
Hence $0 < w(\alpha) \leq w(\alpha \upharpoonright_{\{p,p+1\}}) = g(0,0)g(1,0) = 0$. Contradiction.
So if $g(1,0) = 0$ then $w = w^{[1]}$ since we can never see a constant that behaves in a new way.
From now on assume $\mathbf{g(1,0) > 0}$.
Let $q \geq 2$ be minimal such that $g(q,0) = 0$. We want to show that such a q can't exist;

assuming $g(1,0) > 0$ of course.
Now assume such a q exists. Then by its minimality $w(ID_q) = \prod_{i=0}^{q-1} g(i,0) > 0$.
Now extending ID_q by one constant to an element of SD_{q+1} implies that $g(q,1) > 0$ since $w(ID_q) > 0$ and $w(ID_{q+1}) = 0$ and of course $w(ID_q) = \sum_{\delta \in SD_{q+1},\, \delta \vDash ID_q} w(\delta)$ which in turn equals $q \cdot g(q,1) \prod_{i=0}^{q-1} g(i,0)$.
Let

$$\gamma' := \bigwedge_{i=2}^{q} R_1 a_i \dots a_i \wedge \bigwedge_{i_1 \in \{1,q+1\},\dots,i_{r_0} \in \{1,q+1\}} R_1 a_{i_1} \dots, a_{i_{r_0}} \wedge \nu$$

where ν decides every literal negative to make γ' a state description on $q+1$. Then $\gamma' \vDash ID_q$ and $\gamma' \in SD_{q+1}^q$ with $S_1(\gamma') = \{1, q+1\}$. Now put $\gamma := \gamma'(a_q \mapsto a_{q+1}, a_{q+1} \mapsto a_q)$. We have $\gamma \in SD_{q+1}^q$ with $S_1(\gamma) = \{1, q\}$ and $\gamma \vDash ID_{q-1}$.
By **(CX)** we find $w(\gamma') = w(\gamma)$ and hence

$$\prod_{i=0}^{q-2} g(i,0) g(q-1,0) g(q,1) = \prod_{i=0}^{q-2} g(i,0) g(q-1,1) g(q,0) \quad .$$

But since $g(q,0) = 0$ and q is minimal in that respect we now find $0 = g(q,1)$. However as seen above $g(q,1) > 0$. Contradiction. So $g(p,0) > 0$ for all p.
Let R be a relation symbol with arity $d \geq 3$. Let us consider any $\beta \in SD_p$. Then there exists an $\alpha \in SD_{p+1}^{p+1}$ extending β.[4] One such α is for instance

$$\alpha := \beta \wedge \bigwedge_{i=1}^{p+1} R a_{p+1} a_i a_i \dots a_i \wedge \bigwedge_{\substack{1 \leq i_1,\dots,i_{d-1} \leq p+1 \\ |\{i_1,\dots,i_{d-1}\}| > 1}} \neg R a_{p+1} a_{i_1} a_{i_2} \dots a_{i_{d-1}} \wedge \sigma \tag{13.4}$$

where σ decides all other literals negative that we have to consider to make α a state description on $p+1$. No two constants look the same over α since for all $1 \leq i < k \leq p+1$

$$\alpha \models R a_{p+1} a_i \dots a_i \wedge \neg R a_{p+1} a_k a_i \dots a_i \quad .$$

So by the definition of α and **(JSP)** $w(\alpha) = w(\beta) g(p,0)$.
For $\alpha(a_1 \mapsto a_{p+1}, a_{p+1} \mapsto a_1)$ we have

$$\alpha(a_1 \mapsto a_{p+1}, a_{p+1} \mapsto a_1) \vDash \bigwedge_{i=1}^{p+1} R a_1 a_i a_i \dots a_i \wedge \bigwedge_{\substack{1 \leq i_1,\dots,i_{d-1} \leq p+1 \\ |\{i_1,\dots,i_{d-1}\}| > 1}} \neg R a_1 a_{i_1} a_{i_2} \dots a_{i_{d-1}} \quad .$$

[4] We here give the in Lemma 34 see page 111 promised state description

Hence for all $1 \leq i < k \leq r \leq p$ we have

$$\alpha(a_1 \mapsto a_{p+1}, a_{p+1} \mapsto a_1) \restriction_r \models Ra_1 a_i a_i \dots a_i \wedge \neg Ra_1 a_k a_i \dots a_i \ .$$

So for all r $S(\alpha(a_1 \mapsto a_{p+1}, a_{p+1} \mapsto a_1) \restriction_r)$ consists only of singletons. By **(CX)** we get

$$w(\alpha) = w(\alpha(a_1 \mapsto a_{p+1}, a_{p+1} \mapsto a_1)) = \prod_{i=0}^{p} g(i,0) = w(\beta) g(p,0) \ . \qquad (13.5)$$

Hence from equation (13.5) we find $w(\beta) = \prod_{i=0}^{p-1} g(i,0)$ for all $\beta \in SD_p$ since none of the $g(i,0)$ nor $g(p,0)$ is zero.
But this means that $w(\beta) = w(\delta)$ for all $\delta \in SD_p$, since β was completely arbitrary. And hence $w = w_\infty$.
Clearly $w^{[1]}$ and w_∞ satisfy **(JSP)** and **(CX)**. □

13.3 (JSP) and Equality

Some of the in this section presented ideas can already be found in Kingman's work (see for instance [13]).

A probability function w on $QFS=$ is said to satisfy **(JSP)** if and only if for all p and all $\tau \in T_p$ with $w(\tau \restriction_{p-1}) > 0$ $w(\tau | \tau \restriction_{p-1})$ only depends on $w(\tau \restriction_{p-1})$ and on the number of $1 \leq i \leq p-1$ such that $\tau \vdash a_p \equiv a_i$.

Theorem 21. *The probability functions on $QFS=$ satisfying **(CX)** and **(JSP)** are characterized by one parameter $\lambda \in [0, +\infty] \subset \mathbb{R} \cup \{+\infty\}$.*
The probability of a table $\tau \in T_p^t$ with spectrum $\langle x_1, \dots, x_t \rangle$ is given by

$$w_\lambda(\tau) := \prod_{i=1}^{t} \prod_{j=0}^{x_i - 1} g(j + \sum_{r=1}^{i-1} x_r, j) = \lambda^t \cdot \left(\prod_{i=1}^{p-1} \frac{1}{i+\lambda} \right) \left(\prod_{j=1}^{t} (x_j - 1)! \right)$$

where $g(r,s) := \frac{s}{r+\lambda}$ for $r \geq s > 0$ and $g(r,0) := \frac{\lambda}{r+\lambda}$ for all $r > 0$ and $g(0,0) = 1$.

We make the usual conventions for $n \in \mathbb{N}$ we set $\frac{n}{\infty} = 0$, $n + \infty = \infty$ and $\frac{\infty}{\infty} = 1$.
Strictly speaking $g(r,s)$ depends on λ. Since we are not going to change this parameter once it has been chosen we suppress it in the notation.

Proof. We will first show that the above defined w_λ are probability functions that satisfy **(JSP)** and **(CX)**. Then we'll prove that any probability function on $QFS{=}$ satisfying these two rationales is of this form for some $0 \le \lambda \le \infty$.
Let $\tau \in T_p^t$ have spectrum $\langle x_1, \dots, x_t\rangle$ then since $\sum_{i=1}^t x_i = p$

$$g(p,0) + \sum_{i=1}^{t} g(p,x_i) = \frac{\lambda}{p+\lambda} + \sum_{i=1}^{t} \frac{x_i}{p+\lambda} = \frac{\lambda}{p+\lambda} + \frac{p}{p+\lambda} = 1 \ . \qquad (13.6)$$

Furthermore $w_\lambda : T \to [0,1]$ since all g are between zero and one. On top of that $w_\lambda(a_1 \equiv a_1) = g(0,0) = 1$. And for $\tau \in T_p^t$ with spectrum as above

$$\sum_{\substack{\tau' \in T_{p+1} \\ \tau' \models \tau}} w_\lambda(\tau') = \prod_{i=1}^{t} \prod_{j=0}^{x_i - 1} g\Big(j + \sum_{r=1}^{i-1} x_r, j\Big) \cdot \Big(g(p,0) + \sum_{i=1}^{t} g(p,x_i)\Big) = w_\lambda(\tau) \ . \qquad (13.7)$$

Hence the w_λ define a probability function on the set of tables T and hence induce a unique probability function on $QFS{=}$ (see Lemma 7 on page 92).
Note that (13.7) has to hold for any probability function on the tables satisfying **(JSP)**. However for other probability functions the values for $g(\cdot,\cdot)$ might be different. What has to hold in any case is that

$$g(p,0) + \sum_{i=1}^{t} g(p,x_i) = 1 \qquad (13.8)$$

since we want want to define a probability function.
Next we have to see that the w_λ satisfy **(CX)**. This follows directly from the definition of $w_\lambda(\tau)$ since it is given in terms of the spectrum of τ.
Clearly every w_λ satisfies **(JSP)** by its very definition.
We have now finished showing that w_λ are probability functions that satisfy **(JSP)** and **(CX)**. We now prove that every probability function satisfying these principles is such a w_λ.
Since we want to construct a probability function we have to have that $g(0,0) = 1 = w_\lambda(a_1 \equiv a_1)$.
Furthermore **(JSP)** implies $g(1,0) + g(1,1) = 1$ what follows from extending $a_1 \equiv a_1$ to T_2. This allows us to define λ implicitly by $g(1,0) = \frac{\lambda}{1+\lambda}$ and $g(1,1) = \frac{1}{1+\lambda}$.
The proof proceeds by showing on induction on n that there can be at most one probability function given by the $g(n,\cdot)$ consistent with the choice of λ. Then, as seen above,

the so defined w_λ satisfy all required conditions.
Before giving the general induction step we calculate $g(2,0), g(2,1)$ and $g(2,2)$ as an example.
By **(CX)** we find $w_\lambda(a_1 \equiv a_3 \wedge \neg a_2 \equiv a_3) = w_\lambda(a_1 \equiv a_2 \wedge \neg a_2 \equiv a_3)$ and hence

$$g(0,0)g(1,0)g(2,1) = g(0,0)g(1,1)g(2,0)$$

or equivalently

$$g(0,0)g(1,0)g(2,1) - g(0,0)g(1,1)g(2,0) = 0 \ . \tag{13.9}$$

By looking at the extension of $a_1 \equiv a_2$ and of $\neg a_1 \equiv a_2$ to T_3 we find by using (13.8)

$$g(2,0) + g(2,2) = 1 \quad \text{and} \quad g(2,0) + 2g(2,1) = 1 \ . \tag{13.10}$$

Solving this system of three linear inhomogeneous equations in (13.9) and (13.10) with unknowns $g(2,0), g(2,1)$ and $g(2,2)$ leads to the unique solutions $g(2,0) = \frac{\lambda}{2+\lambda}$, $g(2,1) = \frac{1}{2+\lambda}$ and $g(2,2) = \frac{2}{2+\lambda}$.
Note that even for $\lambda = \infty$ we find the correct solutions as $g(2,2) = g(2,1) = 0$ and $g(2,0) = 1$. For $\lambda = 0 = g(1,0)$ we find $g(2,0) = 0, g(2,2) = 1$ and $g(2,1) = 1/2$. The very last value is not what we advertised it to be. As we are going to see that does not matter.
Now by the induction hypothesis we can assume that all the values for $g(r,\cdot)$ with $r \leq n$ found so far are strictly greater than zero for $\lambda \notin \{0,\infty\}$ and $g(r,s)$ takes values as stated in the theorem. The cases $\lambda \in \{0,\infty\}$ are treated separately at the end.
We find by **(CX)** and extending the unique table τ in T_n^n to T_{n+2}^{n+1} that

$$g(n,0)g(n+1,1) = g(n,1)g(n+1,0) \tag{13.11}$$

and by using (13.8) we find by extending $\tau \in T_{n+1}^1$ to T_{n+2}

$$g(n+1,0) + g(n+1,n+1) = 1 \ . \tag{13.12}$$

By extending a table with spectrum $\langle n+1-k, 1@k\rangle$ to T_{n+2} we find for $1 \leq k \leq n$

again via (13.8)

$$g(n+1,0)+g(n+1,n+1-k)+kg(n+1,1)=1\ . \tag{13.13}$$

These are $2+n$ linear equations in the $n+2$ unknowns $g(n+1,n+1-k)$ for $0\leq k\leq n+1$.
Let's first set $k=n$. Then we get three linear equations in the unknowns $g(n+1,0)$, $g(n+1,1)$ and $g(n+1,n+1)$ which we can solve with unique solutions.
We find $g(n+1,0)=\frac{\lambda}{n+1+\lambda}$, $g(n+1,1)=\frac{1}{n+1+\lambda}$ and $g(n+1,n+1)=\frac{n+1}{n+1+\lambda}$ by using the induction hypothesis to find the values of $g(n,0)$ and $g(n,1)$.
Now using equation (13.13) for $k\in\{1,2,\ldots,n-1\}$ we can calculate directly

$$g(n+1,n+1-k)=1-\frac{\lambda}{n+1+\lambda}-k\frac{1}{n+1+\lambda}=\frac{n+1-k}{n+1+\lambda}\ .$$

Hence once we have chosen our λ there's only one probability function fitting the bill, namely w_λ.
Now assume we chose λ to be zero. That implies $g(1,1)=1$ and hence $w_0(a_1\equiv a_2)=1$. **(CX)** then implies that every two constants have to be the same with probability one and hence w_0 is $1-$heterogeneous. So $g(n,n)=1$ and $g(n,0)=0$ for all n. For $1\leq i<n$ the values of $g(n,i)$ don't matter here.
The last remaining case is $\lambda=\infty$. Hence $g(1,0)=1$ and so $w_\infty(\neg a_1\equiv a_2)=1$. **(CX)** then implies that no two constants can ever be the same and hence w_∞ is given by the prior that puts all measure on $u^{p_\infty}_{=}$, implying $g(n,0)=1$ for all n. The values of $g(n,i)$ for $i>0$ are inconsequential in that case. □

Remark 15. For any relational language L the w_λ above induce a unique probability function w^λ on $QFSL^=$ satisfying **(SX=)**. It would be of considerable structural beauty if there was closed formula for $w^\lambda(\alpha)$ using only generic properties of $\alpha\in SD_p(L)$ and w_λ.
For $\tau\in T^t_p$ and $\alpha\in SD^s_p$ we find if $\alpha\wedge\tau$ is consistent that $w^\lambda(\tau\wedge\alpha)|SD_t|=w_\lambda(\tau)$ (see Corollary 6 on page 91) and of course $w^\lambda(\alpha)=\sum_{\tau\in T_p,\ S(\alpha)\leq S(\tau)}w^\lambda(\tau\wedge\alpha)$.

So overall

$$
\begin{aligned}
w^{\lambda}(\alpha) &= \sum_{\substack{\tau \in T_p^{\geq s}, \\ S(\alpha) \leq S(\tau)}} \frac{w_\lambda(\tau)}{|SD_{|S(\tau)|}(L)|} = \sum_{\substack{E \in ER_p^{\geq s} \\ S(\alpha) \leq E}} \frac{w_\lambda(\tau^E)}{|SD_{|E|}(L)|} \\
&= \left(\prod_{i=1}^{p-1} \frac{1}{i+\lambda} \right) \cdot \left(\sum_{g=0}^{p-s} \frac{\lambda^{s+g}}{|SD_{s+g}|} \sum_{\substack{E \in ER_p^{s+g} \\ S(\alpha) \leq S(E)}} \prod_{j=1}^{s+g} (|E_j| - 1)! \right)
\end{aligned}
$$

where τ^E is the unique table such that $S(\tau) = E$ (see remark 9 on page 92).
We can't expect too nice a closed formula as for not purely unary languages L only $w^{[1]}$ and w_∞ satisfy **(JSP)** and **(CX)** on L. So $w^\lambda \upharpoonright_L$ in general doesn't satisfy **(JSP)**.

Chapter 14

Principles of Conformity

14.1 Preparations and Examples

The idea of the principles of conformity can be traced back to [22] where purely binary languages were considered. A more general approach was explored in [17]. The here presented material is the to date most general.

Here we think of state descriptions as elements of SD^{01}. As mentioned before this does not make a mathematical difference but it here helps to keep the notational nightmare from getting too bad.
Until further notice we will assume that we are given a fixed not purely unary language L. We want to encourage the reader to have another look at the canonical representation chapter (page 32) as the there presented ideas will play a major role here.

In the following we will concern ourselves with certain subformulae of state descriptions and show that **(SX)** implies that they have the same probability. This will lead to another notion of language invariance.
Before taking the long and stony paths to the lofty notion of $\#-$language invariance we give some more down to earth examples.

Example 8. Let H be a $2-$dimensional zero-one array on p. Let

$$\alpha := \bigwedge_{i_1,i_2=1}^{p} R^{(H)_{i_1 i_2}}\ a_{i_1} a_{i_2}$$

$$\beta := \bigwedge_{i_1,i_2=1}^{p} P^{(H)_{i_1 i_2}}\ a_1 a_{i_1} a_{i_2}, \qquad \gamma := \bigwedge_{i_1,i_2=1}^{p} Q^{(H)_{i_1 i_2}}\ a_{i_1} a_{i_1} a_4 a_{i_2} a_{i_1}$$

where R is binary, P is ternary and Q is a $5-$ary relation symbol.
All those examples have in common that if one considers a $\mathcal{G} \in SD_p^{01}(L)$ over a language L that contains P, Q and R then we can always rediscover H in a certain way as a subarray in the array of $\mathcal{G}$ representing R respectively P, Q.
Next we use the examples to introduce some terminology which will be defined precisely shortly.
Let d be the arity of the relation symbol (in the examples above $2, 3$ respectively 5). We want to talk about how the given array H of dimension 2 (later in the general setting of dimension d' with $d' \leq d$) is placed inside an array representing a state description over a $d-$ary relation symbol.
For the first example α the given H stays untouched.
For β we have a ternary relation symbol; $d = 3$. Here H is what one could call the first slice of an array representing a state description P. Technically we inserted one *dummy* 1 to make β a well formed formula.
For γ we inserted one dummy 4 at position 3 and copied the *running index* i_1 to three positions.
In general we take $1 \leq i_1, \ldots, i_{d'} \leq p$ and somehow make this into a tuple with $d \geq d'$ entries. We do this by inserting dummies and by copying running indices.
We will use a first function n_1 to specify where to we copy the running indices and a second function n_2 to specify where which dummy will appear. For instance for $1 \leq k \leq p$ $n_2(k) = \{j, s\}$ will mean that the dummy k will be at the $j-$th position and the $s-$th position and nowhere else.
n_1 and n_2 take values in the powerset, denoted by $\mathcal{P}$, of $\{1, \ldots, d\}$. For $1 \leq u \leq d'$ we will have that $n_1(u)$ can never be empty as we want to "preserve the dimension" of our given array in contrast to $n_2(\cdot)$ which might very well be empty.
We will bundle these functions n_1 and n_2 into one *inserting* function $\iota^{\#}$.
The whole process starting with a given zero-one array and ending up with a first order formula (in the above example α, β, γ) given by an inserting function $\iota^{\#}$ will be referred to as a *decent imbedding* ι.

We will show that any probability function w on a language L containing R, P, Q that satisfies **(SX)** has to give the same probability to α, β and γ.
The idea is to show that $w(\alpha), w(\beta), w(\gamma)$ only depend on d' and the spectrum of H, which we will define shortly. We will show this by counting the number of state

descriptions σ in $SD_q(L)$[1] such that $\sigma \vDash \alpha$ where the spectrum of σ is a given $\vec{x} \in X_q$. As it will turn out the number of such σ only depends on d' and the spectrum of H. Having proved this we have

$$w(\alpha) = \sum_{\vec{x} \in X_q} \sum_{\substack{\sigma \in SD_q \\ \sigma \vDash \alpha \\ \overrightarrow{S(\sigma)} = \vec{x}}} w(\vec{x})$$

and hence $w(\alpha) = w(\beta) = w(\gamma)$.

Recall from definition 5 on page 13 that $\mathbb{N} := \{1, 2, \ldots\}$ and fix a number p. All our given arrays will be on p.

Definition 54. For a fixed $d' \in \mathbb{N}$ let $n_1 : \{1, \ldots, d'\} \to \mathcal{P}(\{1, \ldots, d\}) \setminus \emptyset$ and $n_2 : \mathbb{N} \to \mathcal{P}(\{1, \ldots, d\})$ be functions.
We call $\langle n_1, n_2 \rangle$ a *normal partition of* d if and only if

- $n_1(u) \cap n_1(v) = \emptyset$ for all $1 \leq u < v \leq d'$,
- $n_2(r) \cap n_2(s) = \emptyset$ for all different $r, s \in \mathbb{N}$,
- $n_1(u) \cap n_2(r) = \emptyset$ for all $1 \leq u \leq d'$ and all $r \in \mathbb{N}$ and finally
- $\{1 \ldots, d\} = \bigcup_{u=1}^{d'} n_1(u) \ \cup \ \bigcup_{r \in \mathbb{N}} n_2(r)$.

We will need to talk about the greatest dummy of n_2, we hence put $N := \max\{r \in \mathbb{N} | n_2(r) \neq \emptyset\}$ if n_2 inserts at least one dummy and $N := 0$ otherwise.
We call a map $\iota^{\#} : \{1, \ldots, p\}^{d'} \to \{1, \ldots, \max\{p, N\}\}^d$ with $\iota^{\#}(\langle i_1, \ldots, i_{d'} \rangle) = \langle m_1, \ldots, m_d \rangle$ an *inserting map* with respect to a normal partition $\langle n_1, n_2 \rangle$ if and only if for all $1 \leq t \leq d$, $1 \leq u \leq d'$ and all $r \in \mathbb{N}$

- if $t \in n_1(u)$ then $m_t = i_u$ and
- if $t \in n_2(r)$ then $m_t = r$.

In the example from above we have $d' = 2$.
For α we have that $n_1(1) = \{1\}, \ n_1(2) = \{2\}$ and $n_2(r) = \emptyset$ for all $r \in \mathbb{N}$ and so $N = 0$.
For β we have $n_2(1) = \{1\}, \ n_1(1) = \{2\}, n_1(2) = \{3\}$ and $n_2(r) = \emptyset$ for $r \geq 2$,

[1] for any big enough q

hence $N = 1$.
For the third example γ we have $n_1(1) = \{1, 2, 5\},\ n_1(2) = \{4\}$, $n_2(4) = \{3\}$ and $n_2(r) = \emptyset$ for all other $r \in \mathbb{N}$ so $N = 4$.

Definition 55. Let $L_d \neq \emptyset$,[2] $d' \leq d$ and let H be a zero-one array of dimension d' on p. Let G be an empty zero-one array[3] of dimension d on $\max\{p, N\}$. Let $\iota^{\#}$ be an inserting function with respect to some normal partition of d. Now we fix some entries in G via:

$$(G)_{\iota^{\#}(i_1,\ldots,i_{d'})} := (H)_{i_1,\ldots,i_{d'}} \quad .$$

We call this half completed array $\iota(H)$ and ι a *decent imbedding*. Hence for all $1 \leq j_1, \ldots, j_d \leq \max\{p, N\}$ $(G)_{j_1,\ldots,j_d}$ is undetermined if and only if $\langle j_1, \ldots, j_d\rangle \notin Im(\iota^{\#})$.
So $\iota(H)$ is an incomplete array that upon completion represents a state description on $\max\{p, N\}$ on a language containing one $d-$ary relation symbol. For the time being however we will not complete the array and so consider $\iota(H)$ to be a conjunction of literals.[4] If there is more than one relation symbol in L_d then there is a choice here as to over which relation symbol the conjunction is meant to be. As we are going to assume **(SX)** later on and hence **(PX)**, this choice will be inconsequential.

From now we **will always assume** that ι is a decent imbedding and $\iota^{\#}$ is an inserting function with respect to some normal partition. So for a probability function w on our language L $w(\iota(H))$ is well defined since $\iota(H)$ is well formed formula over L.
If L_{d_1} and L_{d_2} are both not-empty, then we explicitly allow that two decent imbeddings ι_1, ι_2 are with respect to different d_1, d_2. So $\iota_1(H), \iota_2(H)$ are well formed formula over L_{d_1}, L_{d_2} respectively.

14.2 A first Conformity Principle

We now propose a first version of a principle based on symmetry considerations. For a fixed zero-one array H, a fixed probability function w on our given language L we have the following

Principle of General Conformity - (PGC)
The probability of $w(\iota(H))$ does not depend on ι.

[2] We want to recall that L_d is the $d-$ary fragment of L.
[3] That is we think of the entries in G as yet undetermined.
[4] If $d = d'$ then $\iota(H)$ is a complete array.

So in the example if P, Q, R are all in L and w is a probability function on L satisfying **(PGC)**, then $w(\alpha) = w(\beta) = w(\gamma)$.
Before proceeding with the next theorem we need some useful observations.
Note that for a language L' containing only one $d'-$ary relation symbol H represents an element of $SD_p(L')$. From now on L' will always stand for this language. In this sense $S(H)$ is well defined and we can hence talk about the spectrum of $S(H)$.
We will use L' mainly to help counting the number of certain conjunctions and to define S of certain arrays as we just did when we defined $S(H)$.

Definition 56. Let $\beta \in QFSL$ be a satisfiable conjunction of literals and $E \in ER_q$ with $q \geq const(\beta)$. We define $vip_E(\beta)$ to be the result of replacing for all $1 \leq i \leq q$ a_i by $a_{vip_E(i)}$ throughout β.
Should a literal λ appear more than once in $vip_E(\beta)$ then we delete all but one appearance of λ.
If $\vec{m} := \langle m_1, \ldots, m_s \rangle$ is a tuple of numbers such that all entries are between one and q then let $vip_E(\vec{m}) := \langle vip_E(m_1), \ldots, vip_E(m_s) \rangle$.
If $\mathcal{G} \in SD_p^{01}(L)$ is a tuple of zero-one arrays representing β[5] and if $vip_E(\beta)$ is satisfiable then let $vip_E(\mathcal{G})$ be the tuple of arrays that represent $vip_E(\beta)$.

In general $vip_E(\beta)$ is not satisfiable. For example for $\beta = Ra_1 \wedge \neg Ra_2 \wedge \neg Ra_3$ and $E = \{1, 2, 3\}$. Then $vip_E(\beta) = Ra_1 \wedge \neg Ra_1$ as we deleted one occurrence of $\neg Ra_1$.
We would like to use $\iota(H)/E$ here. However if one of the inserted dummies is not a representative of E then $\iota(H)/E$ is the empty formula. That is why we introduce $vip_E(\psi)$.

Definition 57. Let $q \geq p,\ F, G \in ER_q$ and $E \in ER_p$. Then we use $E \leq_p F$ as shorthand for $E \leq F \upharpoonright_p$, we say that F *is a $p-$refinement of* E. We will use the abbreviation $E \leq_p F \leq G$ to denote that $E \leq_p F$ and $F \leq G$. Furthermore let $E =_p F$ stand for $E = F \upharpoonright_p$ and $E <_p F$ for $E < F \upharpoonright_p$.

Let $E \in ER_q$ and let H be zero-one array of dimension d' on p such that E is a $p-$refinement of $S(H)$. Note for later use that $|E \upharpoonright_p|^{d'}$ equals the number of entries of $H/E = H/(E \upharpoonright_p)$ and that the array H/E represents an element of $SD_{|E \upharpoonright_p|}(L')$.

Lemma 38. Suppose γ is of the form $\iota(H)$ for some array H as above. If $S(H) \leq_p E$ for $E \in ER_{\geq \max\{p,N\}}$ then $vip_E(\gamma)$ is satisfiable.

[5] Some arrays of $\mathcal{G}$ may be partially or completely empty here.

Proof. We will here show that for $x \neq y$ with $vip_E(x) = vip_E(y)$ and such that $\langle i_1, \ldots, i_l, x@b, i_{l+1}, \ldots, i_{d-b}\rangle, \langle i_1, \ldots, i_l, y@b, i_{l+1}, \ldots, i_{d-b}\rangle \in Im(\iota^{\#})$[6] that
$(\iota(H))_{\langle i_1,\ldots,i_l,x@b,i_{l+1},\ldots,i_{d-b}\rangle} = (\iota(H))_{\langle i_1,\ldots,i_l,y@b,i_{l+1},\ldots,i_{d-b}\rangle}$.
Since $vip_E(x) = vip_E(y)$ x and y are in the same class of E.
Note that since $x \neq y$ we have $l+1, \ldots, l+b \in Im(n_1)$. That is we can't have a dummy here so we have copies of a running index. So $1 \leq x, y \leq p$. And since $S(H) \leq E \restriction_p$ we find $S(H)xy$.
To keep the notation somewhat tractable we now assume that there is *only one* u such that $n_1(u) = \{l+1, \ldots, l+b\}$. In general there are $u_1, \ldots, u_o$ such that $\bigcup_{c=1}^{o} n_1(u_c) = \{l+1, \ldots, l+b\}$. Also let $\langle r_1, \ldots, r_{d'}\rangle, \langle s_1, \ldots, s_{d'}\rangle$ be such that $\iota^{\#}(\langle r_1, \ldots, r_{d'}\rangle) = \langle i_1, \ldots, i_l, x@b, i_{l+1}, \ldots, i_{d-b}\rangle$ and such that
$\iota^{\#}(\langle s_1, \ldots, s_{d'}\rangle) = \langle i_1, \ldots, i_l, y@b, i_{l+1}, \ldots, i_{d-b}\rangle$.
Then $\langle r_1, \ldots, r_{u-1}, x, r_{u+1} \ldots, r_{d'}\rangle = \langle s_1, \ldots, s_{u-1}, x, s_{u+1} \ldots, s_{d'}\rangle$.
Since x and y are equivalent over $S(H)$ we have

$$(H)_{r_1,\ldots,r_{u-1},x,r_{u+1}\ldots,r_{d'}} = (H)_{s_1,\ldots,s_{u-1},y,s_{u+1}\ldots,s_{d'}}$$

and so as required

$$(\iota(H))_{\iota^{\#}(r_1,\ldots,r_{u-1},x,r_{u+1}\ldots,r_{d'})} = (\iota(H))_{\iota^{\#}(s_1,\ldots,s_{u-1},y,s_{u+1}\ldots,s_{d'})} \quad .$$

□

We want to remark that for the unique $U \in ER_q^q$ and $\beta \in QFSL$ with $const(\beta) \leq q$ we have that $vip_U(\beta) = \beta$ since for all $1 \leq i \leq q$ we replace a_i by a_i in β.
Furthermore for $\delta \in SD_p(L)$ and $E \in ER_q$ with $S(\delta) \leq_p E$ we have that $vip_E(\delta) = \delta/E,\ vip_E(\delta) \cdot E \vDash \delta$ and for $\sigma \in SD_{|E|}(L)$ we find $vip_E(\sigma \cdot E) = \sigma$.

Corollary 14. Let H be d'−dimensional zero-one array on p and let ι be a decent imbedding. Let $E \in ER_q$ with $q \geq \max\{N, p\}$ be such that $S(H) \leq_p E$ and identify $\iota(H)$ with first order formula $\alpha \in QFSL$ it represents. Then

$$\{\gamma \in SD_q(L) \mid S(\gamma) = E,\ \gamma \vDash \alpha\} \overset{1:1}{\longleftrightarrow} \{\delta \in SD_{|E|}^{|E|}(L) \mid \delta \vDash vip_E(\alpha)\} \quad . \qquad (14.1)$$

Proof. We already know from the canonical representation results that

$$\{\gamma \in SD_q(L) \mid S(\gamma) = E\} \overset{1:1}{\longleftrightarrow} SD_{|E|}^{|E|}(L) \quad .$$

[6] We assume here w.l.o.g. for notational convenience that the x and y appear next to each other.

If γ is an element of the set on the left in (14.1) then $\gamma/E = vip_E(\gamma)$ and hence $\gamma/E \vDash vip_E(\alpha)$.
If δ is an element of the set on the right in (14.1) then $\delta \cdot E \vDash \alpha$. □

Lemma 39. Let H be $d'-$dimensional zero-one array on p and let ι be a decent imbedding. Let $E \in ER_q$ with $q \geq \max\{N, p\}$ be such that $S(H) \leq_p E$. Then the number of entries in $vip_E(\iota(H))$ equals $|E \upharpoonright_p|^{d'}$ and does not depend on ι.

Proof. Consider the following diagram [explanations coming up shortly]

$$\begin{array}{ccc} H & \xrightarrow{\iota} & \iota(H) \\ {\scriptstyle vip_E}\downarrow & & \downarrow{\scriptstyle vip_E} \\ vip_E(H) & \xrightarrow[\iota_E]{} & vip_E(\iota(H)) \end{array}$$

ι_E is given in terms of the inserting function $\iota_E^{\#}$ which is defined as follows. $\iota_E^{\#}$ is with respect to the same normal partition $\langle n_1, n_2\rangle$ as $\iota^{\#}$.
$\iota_E^{\#}$ is a function from $\{1, \ldots, |E \upharpoonright_p|\}^{d'}$ to $\{1, \ldots, \max\{p, N\}\}^d$. $\iota_E^{\#}(\langle j_1, \ldots, j_{d'}\rangle)$ equals $\langle m_1, \ldots, m_d\rangle$ if and only if for all $1 \leq t \leq d$, $1 \leq u \leq d'$ and all $r \in \mathbb{N}$

- if $t \in n_1(u)$ then $m_t = j_u$ and
- if $t \in n_2(r)$ then $m_t = vip_E(r)$.

So we find for all $1 \leq i_1, \ldots, i_{d'} \leq p$ that

$$vip_E(\iota^{\#}(i_1, \ldots, i_{d'})) = \iota_E^{\#}(vip_E(i_1, \ldots, i_{d'}))$$

and so

$$vip_E(\iota(H)) = \iota_E(vip_E(H)) \enspace .$$

Hence the diagram commutes and ι_E is a decent imbedding. If follows that $vip_E(H)$ has the same number of entries as $(vip_E(\iota(H)))$. Recall that $vip_E(H) = H/E$ represents an element of $SD_{|E\upharpoonright_p|}(L')$ and hence it consists of $|E \upharpoonright_p|^{d'}$ entries. □

14.3 (SX) implies (PGC)

Theorem 22. *(SX) implies the Principle of General Conformity.*

It is clearly enough to prove the following lemma

Lemma 40. Let H be a $d'-$dimensional zero-one array on p and let ι be a decent imbedding. Let w a probability function on L satisfying **(SX)**.
Then $w(\iota(H))$ only depends on d' and the spectrum of $S(H)$.

We want to mention that p can be obtained from the spectrum of $S(H)$ since the spectrum is in X_p.

Proof. Identify for notational convenience $\iota(H)$ with the first order formula $\alpha \in QFSL$ it represents. Our proof proceeds by showing that for all $E \in ER_q$ with $q \geq \max\{N, p\}$[7] that

$$N(\alpha, E, L) := |\{\delta \in SD_q(L) \mid S(\delta) = E,\ \delta \vDash \alpha\}| \tag{14.2}$$

only depends on d', E and $S(H)$. Assume until further notice that we have proved this, then $N(\alpha, E, L)$ is well defined with this notation for any such q

$$w(\alpha) = \sum_{E \in ER_q} N(\alpha, E, L) \ \cdot \ w(\overrightarrow{\mathcal{S}(E)}) \ .$$

Note that for $\sigma \in S^q$ we have $\delta \in N(\alpha, E, L)$ if and only if $\delta(\sigma) \in N(\alpha(\sigma)^8, \sigma(E), L)$. Let H^1, H^2 be zero-one arrays on p of dimension d' that have the same spectrum and identify β with $\iota(H^1)$ and γ with $\iota(H^2)$. Then there is some $\tau \in S^p$ such that $S(H^1) = \tau(S(H^2))$ and hence

$$\begin{aligned} w(\beta) &= \sum_{E \in ER_q} N(\beta, E, L) \ \cdot \ w(\overrightarrow{\mathcal{S}(E)}) \\ &= \sum_{\tau_q(E) \in ER_q} N(\gamma, \tau_q(E), L) \ \cdot \ w(\overrightarrow{\mathcal{S}(E)}) \\ &= w(\gamma) \ . \end{aligned}$$

Let H be a zero-one array on p of dimension d' and let ι_1, ι_2 be decent imbeddings with greatest dummies N_1, N_2 respectively. Let $q := \max\{N_1, N_2, p\}$ and identify β

[7] Recall that N is biggest dummy inserted by $\iota^{\#}$.

[8] $\alpha(\sigma)$ is obtained from α by replacing every constant a_i in α by $a_{\sigma(i)}$. So far we had only $\alpha(\sigma)$ for $\alpha \in SD_p$ and $\sigma \in S^p$ defined.

with $\iota_1(H)$ and γ with $\iota_2(H)$ then

$$\begin{aligned} w(\beta) &= \sum_{E \in ER_q} N(\beta, E, L) \cdot w(\overrightarrow{\mathcal{S}(E)}) \\ &= \sum_{E \in ER_q} N(\gamma, E, L) \cdot w(\overrightarrow{\mathcal{S}(E)}) \\ &= w(\gamma) \ . \end{aligned}$$

It follows from these two calculations that if $N(\alpha, E, L)$ only depends on d', E and $S(H)$ then $w(\alpha)$ only depends on d' and the spectrum of $S(H)$. It remains to show that $N(\alpha, E, L)$ only depends on d', E and $S(H)$. We will now verify this claim.

We can here assume that E is a $p-$refinement of $S(H)$. If E is not a $p-$refinement of $S(H)$ then $N(\alpha, E, L) = 0$. Since if we can tell two constants apart over a subformula of a state description we can surely tell them apart over any state description containing the subformula.

We proceed by induction on equivalence relations. We start off with $E \in ER_q^{|S(H)|}$ such that $E \upharpoonleft_p = S(H)$ and then proceed to ever finer relations. For the induction step we assume the induction hypothesis for all equivalence relations $F \in ER_q$ such that $S(H) \leq_p F < E$.

For the **base case** we have $E \in ER_q^{|S(H)|}$ and $E \upharpoonleft_p = S(H)$ hence $|E| = |E \upharpoonleft_p|$.

We will show that

$$N(\alpha, E, L) = \frac{|SD_{|E|}(L)|}{|SD_{|E|}(L')|} \ . \tag{14.3}$$

By Corollary 14 we have $N(\alpha, E, L) = |\{\delta \in SD_{|E|}^{|E|}(L) \mid \delta \vDash vip_E(\alpha)\}|$ and by Lemma 39 $vip_E(\alpha)$ consists of $|E|^{d'}$ many literals.[9] If $\delta \in SD_{|E|}(L)$ is such that $\delta \vDash vip_E(\alpha)$ then we distinguish $|S(H)|$ constants over δ implying $\delta \in SD_{|E|}^{|E|}(L)$. So (14.3) holds and hence $N(H, E, L)$ only depends on d', E and $S(H)$.

For the **induction step** consider an $E \in ER_q^{>|S(H)|}$ such that $S(H) \leq_p E$. We will show that

$$\sum_{\substack{F \in ER_q \\ S(H) \leq_p F \leq E}} N(\alpha, F, L) = \frac{|SD_{|E|}(L)|}{|SD_{|E \upharpoonleft_p|}(L')|} \ . \tag{14.4}$$

Note that $|\{\delta \in SD_q(L) \mid S(\delta) \leq E\}| = |SD_{|E|}(L)|$ by the canonical representation

[9] Note that $|E|^{d'}$ equals the number of literals of a state description in $SD_{|E|}(L')$.

results. We hence have for $\gamma \in SD_q(L)$ that

$$\gamma \in \bigcup_{\substack{F \in ER_q \\ S(H) \leq_p F \leq E}} \{\delta \in SD_q(L) \mid S(\delta) = F,\ \delta \vDash \alpha\}$$

$$\iff \gamma/E \in \{\sigma \in SD_{|E|}(L) \mid \sigma \vDash vip_E(\alpha)\}$$

since if $\gamma/E \vDash vip_E(\alpha)$ then $S(H) \leq_p S((\gamma/E) \cdot E)$ and clearly $S((\gamma/E) \cdot E) \leq E$. Conversely if $\delta \in SD_q(L)$ with $S(H) \leq_p S(\delta) \leq E$ is such that $\delta \vDash \alpha$ then we have $\delta/E \vDash vip_E(\alpha)$.

Since $vip_E(\alpha)$ consists of $|E \upharpoonright_p|^{d'}$ literals equation (14.4) holds.

Now recall that our induction hypothesis allows us to conclude that for all $F \in ER_q$ with $S(H) \leq_p F < E$ that $N(\alpha, F, L)$ only depends on d', E and $S(H)$. Hence (14.4) allows us to conclude the same for $N(\alpha, E, L)$ since

$$N(\alpha, E, L) = \frac{|SD_{|E|}(L)|}{|SD_{|E\upharpoonright_p|}(L')|} - \sum_{\substack{F \in ER_q \\ S(H) \leq_p F < E}} N(\alpha, F, L)$$

which completes the proof. □

14.4 The General Rationale

We now consider the simultaneous imbedding of n zero-one arrays $H_1, \ldots, H_n$ on p, where H_i is of dimension d'_i.

Again we assume that we are given some polyadic language L.

Recall that we defined $m_s(L)$ to be the number of relation symbols of arity s in L. Let $\vec{d'} := \langle d'_1, \ldots, d'_n \rangle$ and $\vec{d} := \langle d_1, \ldots, d_n \rangle$ be such that

1. $1 \leq d'_i \leq d_i$ for all $1 \leq i \leq n$ and
2. $|\{r \mid d_r = s\}| \leq m_s(L)$.

For all $1 \leq i \leq n$ let $\langle n^i_1, n^i_2 \rangle$ be normal partitions of d_i. As before let N_i be the greatest dummy of n^i_2. We put $N := \max_i\{N_i\}$.

Let $\mathbb{H} := \langle H^1, \ldots, H^n \rangle$ be a tuple of zero-one arrays on p of dimension $d'_1, \ldots, d'_n$ respectively. Let $\mathbb{G} := \langle G^1, \ldots, G^n \rangle$ be a tuple of different empty zero-one arrays on $\max\{N, p\}$ of dimension $d_1, \ldots, d_n$.

For $1 \leq i \leq n$ let ι_i be a decent imbedding of H^i into G^i. We collect the maps ι_i to one comprehensive function J. So $J(\mathbb{H})$ fixes some values of $\mathbb{G}$.
Let L' be the language that satisfies for all $1 \leq i \leq \max_j\{d'_j\}$ that $m_i(L') = |\{r|d'_r = i\}|$ and $m_j(L') = 0$ for all other j.
We chose L' such that $\mathbb{H}$ represents a state description in $SD_p(L')$. This allows us to define $S(\mathbb{H})$ in the usual way. And since now $S(\mathbb{H}) \in ER_p$ we can define the *spectrum* of $\mathbb{H}$ in the usual way as the spectrum of $S(\mathbb{H})$. Furthermore $\mathbb{H}$ consists of $\sum_{i=1}^n p^{d'_i}$ many entries.
Let $E \in ER_q$ with $q \geq \max\{N, p\}$ be such that $S(\mathbb{H}) \leq_p E$ [10] then $vip_E(\mathbb{H}) = \langle vip_E(H^1), \ldots, vip_E(H^n)\rangle$. By applying our previous arguments to all H^i and ι_i it follows that for $E \in ER_q$ with $S(\mathbb{H}) \leq_p E$ that $vip_E(J(\mathbb{H}))$ is consistent and consists of $\sum_{i=1}^n |E \upharpoonright_p|^{d'_i}$ many entries. This number equals the number of literals of a formula in $SD_{|E\upharpoonright_p|}(L')$.
We have after identifying $J(\mathbb{H})$ with first order formula α it represents and for $S(\mathbb{H}) \leq_p E$ that

$$\{\gamma \in SD_q(L) \mid S(\gamma) = E,\ \gamma \models \alpha\} \overset{1:1}{\longleftrightarrow} \{\delta \in SD_{|E|}^{|E|}(L) \mid \delta \models vip_E(\alpha)\} \ .$$

We now introduce the general rationale. For a fixed $\mathbb{H}$ as above, a fixed probability function w on our given language L we have the following

Generalized Principle of Conformity - (GPC)
The probability of $w(J(\mathbb{H}))$ does not depend on the imbedding J.

So the principle implies that the probability of $w(J(\mathbb{H}))$ does not depend on $\vec{d}$ as long as it satisfies the conditions 1 and 2 above. $\vec{d'}$ is not a parameter as we assume that $\mathbb{H}$ is fixed.

14.5 (SX) implies (GPC)

Theorem 23. *(SX) implies the Generalized Principle of Conformity.*

Again we will prove a stronger lemma

Lemma 41. Let $\mathbb{H}$ be an $n-$tuple of $d'_1, \ldots, d'_n-$dimensional zero-one arrays on p and

[10] Then clearly $S(H^i) \leq_p E$ for all $1 \leq i \leq n$.

let J be a decent imbedding as above. Let w a probability function on L satisfying **(SX)**.
Then $w(J(\mathbb{H}))$ only depends on $\vec{d'}$ and the spectrum of $S(\mathbb{H})$.

We again mention that p can be obtained from the spectrum of $S(\mathbb{H})$ since the spectrum is in X_p.

Proof. All the main ideas already appear in the above proof. We will recycle the above proof with the slight adjustments needed here.
Identify for notational convenience $J(\mathbb{H})$ with α. Our proof proceeds by showing that for all $q \geq \max\{N, p\}$ and $E \in ER_q$ that

$$N(\mathbb{H}, E, L) := |\{\delta \in SD_q(L) \mid S(\delta) = E,\ \delta \vDash \alpha\}|$$

only depends on $\vec{d'}$, E and $S(\mathbb{H})$ and $S(\mathbb{H})$ and his hence well defined. Then with this notation

$$w(J(\mathbb{H})) = \sum_{E \in ER_q} N(\mathbb{H}, E, L) \cdot w(\overrightarrow{\mathcal{S}(E)}) \ .$$

The same arguments as above show that if $N(\mathbb{H}, E, L)$ only depends on $\vec{d'}$, E and $S(\mathbb{H})$ then $w(J(\mathbb{H}))$ only depends on $\vec{d'}$ and the spectrum of $S(\mathbb{H})$. It remains to show that $N(\mathbb{H}, E, L)$ only depends on $\vec{d'}$, E and $S(\mathbb{H})$.

As ever we can assume that E is a $p-$refinement of $S(\mathbb{H})$. If E is not a $p-$refinement of $S(\mathbb{H})$ then $N(\mathbb{H}, E, L) = 0$. Since if we can tell two constants apart over a subformula of a state description we can surely tell them apart over any state description containing the subformula.
We proceed by induction on equivalence relations. We start of with $S(\mathbb{H}) =_p E$ and then proceed to ever finer relations. For the induction step we assume the induction hypothesis for all equivalence relations $F \in ER_q$ such that $S(\mathbb{H}) \leq_p F < E$.
For the **base case** we have $E \in ER_q^{|S(\mathbb{H})|}$ and $E \upharpoonright_p = S(\mathbb{H})$. We will show that

$$N(\mathbb{H}, E, L) = \frac{|SD_{|E|}(L)|}{|SD_{|E|}(L')|} \ . \tag{14.5}$$

We have seen above that $N(\mathbb{H}, E, L) = |\{\delta \in SD_{|E|}^{|E|}(L) \mid \delta \vDash vip_E(\alpha)\}|$ and that $vip_E(\alpha)$ consists of $\sum_{i=1}^{n} |E|^{d'_i}$ many literals. If $\delta \in SD_{|E|}(L)$ is such that $\delta \vDash vip_E(\alpha)$ then we distinguish $|E|$ constants over δ implying $\delta \in SD_{|E|}^{|E|}(L)$. So (14.5) holds and hence $N(\mathbb{H}, E, L)$ only depends on $\vec{d'}$, E and $S(\mathbb{H})$.

For the **induction step** consider an $E \in ER_q^{>|S(\mathbb{H})|}$ such that $S(\mathbb{H}) \leq_p E$. We will show that

$$\sum_{\substack{F \in ER_q \\ S(\mathbb{H}) \leq_p F \leq E}} N(\mathbb{H}, F, L) = \frac{|SD_{|E|}(L)|}{|SD_{|E \upharpoonright_p|}(L')|} \quad . \tag{14.6}$$

Note that $|\{\delta \in SD_q(L) \mid S(\delta) \leq E\}| = |SD_{|E|}(L)|$ by the canonical representation results. We hence have for $\gamma \in SD_q(L)$ that

$$\gamma \in \bigcup_{\substack{F \in ER_q \\ S(H) \leq_p F \leq E}} \{\delta \in SD_q(L) \mid S(\delta) = F,\ \delta \vDash \alpha\}$$
$$\Longleftrightarrow \gamma / E \in \{\sigma \in SD_{|E|}(L) \mid \sigma \vDash vip_E(\alpha)\}$$

since if $\gamma/E \vDash vip_E(\alpha)$ then $S(\mathbb{H}) \leq_p S((\gamma/E) \cdot E)$ and clearly $S((\gamma/E) \cdot E) \leq E$. Conversely if $\delta \in SD_q(L)$ with $S(\mathbb{H}) \leq_p S(\delta) \leq E$ such that $\delta \vDash \alpha$ then $\delta/E \vDash vip_E(\alpha)$.
Since $vip_E(\alpha)$ consists of $\sum_{i=1}^{n} |E \upharpoonright_p|^{d_i'}$ literals, equation (14.6) holds.
Now recall that our induction hypothesis allows us to conclude that for all $F \in ER_q$ with $S(\mathbb{H}) \leq_p F < E$ that $N(\mathbb{H}, F, L)$ only depends on $\vec{d'}$, E and $S(\mathbb{H})$. Hence (14.6) allows us to conclude the same for

$$N(\mathbb{H}, E, L) = \frac{|SD_{|E|}(L)|}{|SD_{|E \upharpoonright_p|}(L')|} - \sum_{\substack{F \in ER_q \\ S(H) \leq_p F < E}} N(\mathbb{H}, F, L)$$

which completes the proof. □

Chapter 15

$\#-$Language Invariance

In this chapter we will only consider probability functions satisfying **(SX)**.
Recall that for a language L we defined $r_0(L)$ to be the maximal arity of a relation symbol in L. Let $L, L\#$ be languages in our sense such that for all $1 \leq d \leq r_0(L)$ we have $\sum_{i=d}^{r_0(L)} m_i(L) \leq \sum_{i=d}^{\infty} m_i(L\#)$.
We now investigate if for a given v on L there is a probability function w on $L\#$ such that for a decent imbedding J of L into $L\#$ and every $\varphi \in SD(L)$ we have $v(\varphi) = w(J(\varphi))$.

Definition 58. We call a language $L\#$ an $\#-$*extension* of L if and only if for all $1 \leq d \leq r_0$ we have

$$\sum_{i=d}^{r_0(L)} m_i(L) \leq \sum_{i=d}^{r_0(L\#)} m_i(L\#) \quad .$$

This condition allows us to find a decent imbedding of the state descriptions of L into $L\#$.

Definition 59. A $\#-$*language invariant family* is a family of probability functions $w_{\mathcal{L}}$ such that whenever a language $L\#$ is an $\#-$extension of a language L and J is a decent imbedding of L into $L\#$ then $w_{L\#}(J(\alpha)) = w_L(\alpha)$ for all $\alpha \in SD(L)$.
A probability function w is said to be $\#-$*language invariant* if and only if it's a member of a $\#-$language invariant family.

We only require here $w_{L\#}(J(\alpha)) = w_L(\alpha)$ for all $\alpha \in SD(L)$. We could require that to hold for arbitrary $\psi \in QFSL$, however we defined the imbedding only for complete arrays $\mathbb{H}$ so we would have to define $J(\psi)$. For arbitrary $\psi \in QFSL$ we assume that we can always use the standard trick and express $w_L(\psi)$ as the sum of probabilities of $\beta \in SD_q(L)$ that imply ψ; for big enough q.

The next two lemmata also hold in the language invariant case and the proofs are very similar.

Lemma 42. If for given v on a not purely unary L the $\#-$extension w on $L\#$ exists, then w is unique.

Proof. Let $\alpha \in SD_p(L)$ with spectrum $\vec{x}$ and consider an imbedding J with greatest dummy N and $q \geq \max\{N, p\}$ then define

$$N\#(\vec{x}, \vec{y}) := |\{\beta \in SD_q(L\#) \mid \overrightarrow{\mathcal{S}(\beta)} = \vec{y},\ \beta \vDash J(\alpha)\}| \qquad (15.1)$$

which as seen in the above proofs does not depend on α but only on its spectrum $\vec{x}$.
If such a w exists then $w(1@q)N\#(1@q, 1@q) = v(1@q)$ for every $q \geq N$.
And for general $\vec{x} \in X_q^t$ with $q \geq N$ we find

$$v(\vec{x}) = \sum_{\vec{y} \in X_q^{\geq t}} N\#(\vec{x}, \vec{y}) \cdot w(\vec{y}) \ . \qquad (15.2)$$

This gives a system of inhomogeneous linear equations in the unknowns $w(\vec{y})$. By induction on $i := q - t$[1] we'll see that there can be at most one solution of (15.2).
We have already seen the base case $0 = i = q - t$.
For the induction step assume that for all $\vec{z} \in X_q^{>t}$ there is at most one solution for (15.2).
Now consider for $\vec{x} \in X_q^t$ with $q \geq N$

$$v(\vec{x}) = N\#(\vec{x}, \vec{x})w(\vec{x}) + \sum_{\vec{z} \in X_q^{>t}} N\#(\vec{x}, \vec{z})w(\vec{z}) \ .$$

This is an equation in one unknown $w(\vec{x})$ as we assume that v is given and by the induction hypothesis so are the $w(\vec{z})$. We can of course always solve this equation, but we do want to define a probability function and so we have to require $w(\vec{x}) \geq 0$.
So we know that for all $\alpha \in SD_{\geq N}$ there is at most one solution for $w(\alpha)$. If the solution exists for all $\alpha \in SD_{\geq N}$ then for all $\beta \in SD_{<N}$ $w(\beta)$ can be obtained from the following equation $w(\beta) = \sum_{\alpha \in SD_N, \alpha \vDash \beta} w(\alpha)$.
Hence there can be *at most one extension* w. □

Assume that in the above proof we have found a solution w. Note that it is a priori not obvious that this solution w is a probability function. To see if that is the case we

[1] where $q \geq N$ is fixed

would have to verify that w satisfies the conditions (P1) and (P2), see Definition 12 on page 16 or we'd have to use Lemma 4.

Lemma 43. Let v be $t-$heterogeneous with $t \geq 2$ on a not purely unary language L. Then there is no $\#-$extension w of v to any $L\#$ with $L \neq L\#$.[2]

The idea here is very similar to the one behind Lemma 28 on page 97.

Proof. We assume that such a w exists and derive a contradiction. Let J be a decent imbedding of L into $L\#$.
Then for $\alpha \in SD_{t+1}^{t+1}(L)$ $0 = v(\alpha) = w(J(\alpha))$ has to hold. This of course implies for all $\gamma \in SD^{>t}(L\#)$ that $w(\gamma) = 0$ and hence w is $\leq t-$heterogeneous.
Recall that v is $t-$heterogeneous implies that $\lim_p \sum_{\alpha \in SD_p^t(L)} v(\alpha) = 1$.
For an $\alpha \in SD_p^t(L)$ since w has to be $\leq t-$heterogeneous and for $q \geq \max\{N, p, t\}$

$$v(\alpha) = \sum_{\substack{\beta \in SD_q^t(L\#) \\ \beta \vDash J(\alpha)}} w(\beta) \ .$$

For different $\alpha, \alpha' \in SD_p^t(L)$ the respective $\beta, \beta' \in SD_q^t(L\#)$ can only appear in one sum since they either extend $J(\alpha)$ or $J(\alpha')$ but never both at the same time. So since

$$\sum_{\alpha \in SD_p^t(L)} v(\alpha) = \sum_{\alpha \in SD_p^t(L)} \sum_{\substack{\beta \in SD_q^t(L\#) \\ \beta \vDash J(\alpha)}} w(\beta)$$

we have

$$1 = \lim_{p \to \infty} \sum_{\alpha \in SD_p^t(L)} v(\alpha) = \lim_{p \to \infty} \sum_{\alpha \in SD_p^t(L)} \sum_{\substack{\beta \in SD_p^t(L\#) \\ \beta \vDash J(\alpha)}} w(\beta) \leq \lim_{p \to \infty} \sum_{\beta \in SD_p^t(L\#)} w(\beta) \leq 1 \ .$$

But then the less-or-equal symbols can be replaced by $=$ and it follows that w is $t-$heterogeneous.
Now consider a $\gamma \in SD_u^{t-1}(L)$ with $u := \max\{t, N\}$ then by $t-$heterogeneity of v

$$\lim_{p \to \infty} \sum_{\substack{\alpha \in SD_p^{t-1}(L) \\ \alpha \vDash \gamma}} \sum_{\substack{\beta' \in SD_p^t(L\#) \\ \beta' \vDash J(\alpha)}} w(\beta') \leq \lim_{p \to \infty} \sum_{\alpha \in SD_p^{t-1}(L)} v(\alpha) = 0 \ .$$

[2] We here say that two languages L, L' are different if there exists a d such that L_d and L'_d contain a different number of relation symbols.

Now consider a $\gamma' \in SD_u^t(L\#)$ such that $\gamma' \vDash J(\gamma)$. Since the spectrum of γ' is of length t $w(\gamma')$ can't be zero. Now using that v is $t-$heterogeneous and that $\gamma' \vDash J(\gamma)$ we find

$$\begin{aligned} w(\gamma') &= \lim_{p\to\infty} \sum_{\substack{\beta \in SD_p^t(L\#) \\ \beta \vDash \gamma'}} w(\beta) \\ &\leq \lim_{p\to\infty} \sum_{\substack{\alpha \in SD_p^{t-1}(L) \\ \alpha \vDash \gamma}} \sum_{\substack{\beta' \in SD_p^t(L\#) \\ \beta' \vDash J(\alpha) \\ \beta' \vDash \gamma'}} w(\beta') + \lim_{p\to\infty} \sum_{\substack{\alpha \in SD_p^t(L) \\ \alpha \vDash \gamma}} \sum_{\substack{\beta'' \in SD_p^t(L\#) \\ \beta'' \vDash J(\alpha) \\ \beta'' \vDash \gamma'}} w(\beta'') \\ &= 0 + \quad \lim_{p\to\infty} \sum_{\substack{\alpha \in SD_p^t(L) \\ \alpha \vDash \gamma}} \sum_{\substack{\beta'' \in SD_p^t(L\#) \\ \beta'' \vDash J(\alpha) \\ \beta'' \vDash \gamma'}} w(\beta'') \\ &= 0 \ . \end{aligned}$$

The last sum is zero since there is no such β''.
For any $\alpha \in SD_p^t(L)$ extending γ we have that all $\beta'' \in SD_p(L\#)$ extending $J(\alpha)$ and γ' have a spectrum of length at least $t+1$. Since we can tell t of the first u constants over γ' apart and at least one constant of $a_{u+1}, \ldots, a_p$ is new over α and so we can at least distinguish $t+1$ constants over β''.
So overall $w(\gamma') = 0$, yielding the required contradiction. □

Lemma 44. Let v be a language invariant (in the old sense) probability function on L. Let μ be a de Finetti prior of v.[3] Then for every $\#-$extension $L\#$ of a language L_0 the $\#-$extension $v^\#$ of v_{L_0} exists and $v^\# = \int u_{L\#}^{\vec{p}} d\mu$. Hence v is $\#-$language invariant.

Proof. Since v is language invariant there is a language invariant probability function w' satisfying **(SX)** extending v_{L_0} to $L' := L_0 \cup L\#$ where the union is disjoint, that is $QFSL_0 \cap QFSL\# = \emptyset$.
Since w' satisfies **(SX)** it satisfies the Generalized Principle of Conformity.
Hence we can imbed v_{L_0} on L_0 into w' on L' where the image of imbedding is in $L\#$ by some imbedding J say. We can of course also imbed L_0 into L' with image L_0 by some imbedding I say. We can chose I to be the identity.[4] Since the w' satisfies **(SX)** the probability does not depend on the imbedding.

[3] If L is purely unary then there might be more than one language invariant family v belongs to and hence more than one prior. The following proof does not depend on which prior one chooses here. $v^\#$ will of course be according to the chosen prior.

[4] That is for $\psi \in SD(L_0)$ we have $I(\psi) = \psi$.

Now set $v^{\#} := w' \upharpoonright_{L\#}$ on $L\# \subset L'$. As a restriction of w' $v^{\#}$ satisfies **(SX)**.
This $v^{\#}$ is already what we are after. $v^{\#}$ is a restriction of w' which is language invariant. Hence $v^{\#}$ is given by the same prior as w' which in turn has the same prior as v. For $\alpha \in SD(L_0)$ we hence have $v_{L_0}(\alpha) = w'(I(\alpha)) = w'(J(\alpha)) = v^{\#}(J(\alpha))$.
So $v^{\#}$ is the $\#-$extension of v_{L_0} as required. □

Lemma 45. If v is $\#-$language invariant then v is language invariant in the old sense.

Proof. Let L, L' be given where L is a sublanguage L'. We have to show that $v_{L'}(\alpha) = v_L(\alpha)$ for all $\alpha \in SD(L)$.
Note that L' is a $\#-$extension of L. Hence the $\#-$extension w of v_L into L' exists.
Again we choose our imbedding to be the identity I such that for $\alpha \in SD(L)$ we have $I(\alpha) = \alpha$. Hence $v_L(\alpha) = w(I(\alpha)) = w(\alpha)$.
Putting $v_{L'} := w$ we now find $v_{L'} \upharpoonright_L = w \upharpoonright_L = v$. □

Theorem 24. *For a probability function v satisfying **(SX)** the following are equivalent*

- *v is language invariant*
- *v is $\#-$language invariant.*

Chapter 16

Conclusions

We set out to investigate rational principles for polyadic relational first order languages which has to the best of our knowledge never been done thoroughly. As outlined in the introduction we took the *mathematical* route and considered the consequences of a rational principles. Our focus was mainly on the principle of spectrum exchangeability.
After showing that the principle of spectrum exchangeability implies a set of basic principles of exchangeability such as **(AX)** and **(CX)** we went on and proved various representation results for probability functions satisfying **(SX)**.
Next we investigated what happens if we not only want **(SX)** but also of language invariance to hold. This led, rather surprising to us, to the idea to include the equality symbol $\equiv$ into our language. We now feel that we have a rather good understanding of these phenomena.
We then went on and compared the absolute value of the probability state descriptions where the Paris Conjecture took center stage.
After that we showed that **(JSP)** and **(CX)** is a toxic mix as there are only two trivial probability functions satisfying both principles on a not purely unary language.
In the last chapters we saw that purely combinatoric arguments can lead to interesting results.

Overall we feel that we have shed a considerable amount of light in the up to now uncharted realm of uncertain reasoning/inductive logic with first order polyadic relational languages.
As it happens so often solving a problem leads to a new set of new questions. For instance is there a way to extend the here presented framework to languages that also contain function symbols?

Also the quest for new interesting rational principles never seems to end. We hope however to have demonstrated that the principle of spectrum exchangeability was worth our attention by allowing representation theorems and displaying unforeseen consequences. Who would have thought that the purely on symmetric considerations based principle of spectrum exchangeability implies that $w(\langle 1@p \rangle) \leq w(\vec{x})$ for any $\vec{x} \in X_p$?

Bibliography

[1] Carnap, R., *Logical Foundations of Probability*, University of Chicago Press, Chicago, Routledge & Kegan Paul Ltd., 1950.

[2] Carnap, R., *The Continuum of Inductive Methods*, University of Chicago Press, 1952.

[3] Carnap, R., *Replies ans systematic expositions,* in *The Philosophy of Rudolf Carnap,* ed. Schilpp, P.A., La Salle, Illinois, Open Court, 1963.

[4] Carnap, R., *A basic system of inductive logic,* part 1, in *Studies in Inductive Logic and Probability (Volume I),* Eds. Carnap, R. & Jeffrey, R.C., University of California Press, Berkeley and Los Angeles, 1971, 33-165.

[5] Carnap, R., *A basic system of inductive logic,* in *Studies in Inductive Logic and Probability*, Volume II, ed. R. C. Jeffrey, University of California Press, 1980, 7-155.

[6] Carnap, R. & Stegmüller, W., *Induktive Logik und Wahrscheinlichkeit,* Springer Verlag, Wien, 1959.

[7] De Finetti, B., La prevision: ses lois logiques, ses sources subjetive, *Annales de l'Institut Henri Poincaré*, 1937, **7**:1-68.

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

[9] Hardy, G.H., Littlewood, J.E. & Pólya, G., *Inequalities*, Cambridge University Press, 1967.

[10] Hoover, D.N., *Relations on probability spaces and arrays of random variables,* Preprint, Institute of Advanced Study, Princeton, 1979.

[11] Johnson, W.E., *Probability: The deductive and inductive problems, Mind*, 1932, **41**(164):409-423.

[12] Kemeny, J.G., Carnap's theory of probability and induction, *The Philosophy of Rudolf Carnap*, ed. P.A.Schilpp, La Salle, Illinois, Open Court, 1963, 711-738.

[13] Kingman, J.F.C., *The Mathematics of Genetic Diversity*, Philadelphia, SIAM, 1980.

[14] Krauss, P.H., Representation of symmetric probability models, *Journal of Symbolic Logic*, 1969, **34**(2):183-193.

[15] Landes, J., Paris, J.B. & Vencovská, A., Language Invariance and Spectrum Exchangeability in Inductive Logic, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Proceedings of the 9th European Conference, ECSQARU 2007, Hammamet, Tunisia, Springer LNAI 4724, 151-160

[16] Landes, J., Paris, J.B. & Vencovská, A., A characterization of the Language Invariant Families satisfying Spectrum Exchangeability in Polyadic Inductive Logic, Proceeding of the IPM Logic Conference, Tehran, 2007, to appear in a special edition of the *Annals of Pure and Applied Logic*.

[17] Landes, J., Paris, J.B. & Vencovská, A., The Principle of Conformity and Spectrum Exchangeability, to appear in *Foundations of Formal Sciences VI*, eds. B.Löwe & J-W.Romeijn, *Studies in Logic*, College Publications, 2008.

[18] Landes, J., Paris, J.B. & Vencovská, A., Representation Theorems for probability functions satisfying Spectrum Exchangeability in Inductive Logic, submitted to the *International Journal of Approximate Reasoning*

[19] Landes, J., Paris, J.B. & Vencovská, A., Instantial Relevance in Polyadic Inductive Logic; *Logic and Its Applications Third Indian Conference,* ICLA 2009, Chennai, India, January 7-11, 2009, Proceedings

[20] Landes, J., Paris, J.B. & Vencovská, A., Some Aspects of Polyadic Inductive Logic; *Studia Logica Special Issue on Many-valued Logic and Cognition* Vol. 90, No. 1. (1 October 2008), pp. 3-16.

[21] Nix, C.J., *Probabilistic Induction in the Predicate Calculus*, Doctorial Thesis, Manchester University, Manchester, UK, 2005.
See `http://www.maths.man.ac.uk/~jeff/#students`

[22] Nix, C.J. & Paris, J.B., A note on binary inductive logic, *Journal of Philosophical Logic*, 2007, **36**(6):735-771.

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

[24] Paris, J.B., *Short course on Inductive Logic, JAIST 2007*, available at `http://www.maths.manchester.ac.uk/~jeff/`

[25] Paris, J.B. & Vencovská, A., From unary to binary inductive logic, to appear in the *Proceedings of the Second Indian Conference on Logic and its Applications,* Mumbai, 2007.

[26] Paris, J.B. & Waterhouse, P., Atom exchangeability and instantial relevance, *Journal of Philosophical Logic,* ISSN 0022-3611 (Print) 1573-0433 (Online), DOI 10.1007/s10992-008-9093-3

[27] Paris, J.B. & Vencovská, A., A generalization of Muirhead's inequality, to appear in *Journal of Mathematical Inequalities*

[28] Vencovská, A., Binary Induction and Carnap's Continuum, Proceedings of the 7th Workshop on Uncertainty Processing (WUPES), Mikulov, 2006. See `http://mtr.utia.cas.cz/wupes06/articles/data/vencovska.pdf`

[29] Walley, P., Inferences from Multinomial Data: Learning about a Bag of Marbles, *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 58, No. 1 (1996), pp. 3-57

[30] Zabell, S.L., 1997. The continuum of inductive methods revisited. In: Earman, J., Norton, J.D. (Eds.), The Cosmos of Science, *Pittsburgh-Konstanz Series in the Philosophy and History of Science*, University of Pittsburgh Press/Universitätsverlag Konstanz, pp. 351-385.

[31] Zabell, S.L., *Symmetry and its Discontents*, Cambridge Studies in Probability, Induction and Decision Theory, General Editor B.Skryms, Cambridge University Press, 2005

Appendix A

Standard Theorems

The Disjunctive Normal Form Theorem and the Prenex Normal Form Theorem

Theorem 25. *Let* $\psi \in SL$ *and let* $p := \max\{i \mid a_i \in \psi\}$.[1]
Then there exists a $\chi \in QFFL$ *such that* $const(\chi) \leq p$, χ *is of the form* $\bigvee_i \bigwedge_k \chi_{ik}$ *where all the* χ_{ik} *are literals and for different* s, u $\bigwedge_k \chi_{uk} \wedge \bigwedge_k \chi_{sk}$ *is inconsistent and*

$$\models \psi \leftrightarrow Q_1 x_1 Q_2 x_2 \dots Q_r x_r \bigvee_i \bigwedge_k \chi_{ik} \tag{A.1}$$

where the Q_i *are quantifiers.*

Proof. Omitted. □

The Theorem of Dominated Convergence

Theorem 26. *Let* $f_1, f_2, f_3, \ldots$ *denote a sequence of real-valued positive measurable functions on a measure space* (S, Σ, μ). *Suppose the sequence* f_n *converges pointwise in* S *and assume that there is a function* $g : S \to \mathbb{R}$ *such that for all* $s \in S$ *we have that* $g(s) \geq f_n(s)$ *for all* n. *Assume further more that* $\int_S g(s) d\mu$ *is finite. Then*

$$\int_S \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_S f_n d\mu \quad . \tag{A.2}$$

[1] Here we allow L to be any first order language. So it might contain a symbol for equality and function symbols.

We will use this result in the following form

Corollary 15. Let $\mathbb{P}$ with $\mathbb{P} \in \{\mathbb{H}, \mathbb{H}_t, \mathbb{B}\}$ and suppose $w_1, w_2, w_3, \dots$ is a sequence of probability functions depending on $p \in \mathbb{P}$ and suppose μ is a probability measure on $\mathbb{P}$. Then

$$\int_{\mathbb{P}} \lim_{n\to\infty} w_n(\varphi, \vec{p}) d\mu = \lim_{n\to\infty} \int_{\mathbb{P}} w_n(\varphi, \vec{p}) d\mu \quad . \tag{A.3}$$

Proof. By the theorem it's enough to find a suitable function g. We put $g : SL \times \mathbb{P} \to \{1\}$ such that for all $\varphi \in SL$ we have $g(\varphi, \vec{p}) = 1$. Then $g(\varphi) = 1 \geq w_n(\varphi, \vec{p})$ and

$$\int_{\mathbb{P}} g(\varphi, \vec{p}) d\mu = \int_{\mathbb{P}} d\mu = 1 \quad .$$

□

Glossary

We now list our main notations. The number refers to the page where we introduced the notation.

$A \subset B$ 12

$|\vec{v}|$ 12

$\langle n@t \rangle$ 12

$\mathcal{H}$ 12

$(H_d)_{j_1,j_2,\ldots,j_d}$ 12

$\mathcal{H} \subset \mathcal{G}$ 13

$\mathbb{N}$ 13

$QFFL/QFSL$ 13

$r_0(L)$ 13

L_d 13

$m_d(L)$ 13

AFL 14

$R^0\vec{a}$ 14

$\bigwedge_{\vec{a}\in A} \pm R\vec{a}$ 14

$SD_p(L)$ 14

$\alpha \in SD_p^{01}$ 15

$const(\psi)$ 17

ER_p 21

$E(i \mapsto k, k \mapsto i)$ 21

$|E|$ 21

ER_p^t 21

$E(i)j$ 21

$E\restriction_{i_1,\ldots,i_k}$ 21

$E\restriction_q$ 21

$|E| \leq |D|$ 21

$V(E)$ 21

$v(i)$ 21

$v(\vec{a})$ 21

vip 21

$\mathcal{S}(E)$ 22

$\overrightarrow{\mathcal{S}(E)}$ 22

$\varphi(a_i \mapsto a_j)$ 22

$S(\alpha)$ 22

$\overrightarrow{\mathcal{S}(\alpha)}$ 24

X_p 25

X_p^t 25

$SD_p^{\leq t}$ 25

(SX) 26

(REG) 26

(SN) 27

(WN) 27

(CX) 27

(PX) 27

S^d 28

(AX) 28

$\sigma(\vec{a})$ 28

(VX) 28

$\psi\upharpoonright$ 32

$\frac{\alpha}{E}$ 33

α/E 33

Red 37

$\alpha \cdot E$ 37

$C_d(H, E).$ 42

$D(\mathcal{H}, E)$ 44

$\alpha(\sigma)$ 45

σ_q 45

$\pi(E)ij$ 45

$G(\alpha, \vec{x})$ 45

$N(\vec{x}, \vec{y}, L)$ 46

$d(\vec{y}, L, L')$ 46

$d(t, L, L')$ 46

$w^{[1]}$ 47

$w^{[t]}$ 52

η_t 52

$w^{(l)}$ 54

B_p 60

$S(c)$ 60

$prob(c)$ 60

$Cont(\alpha|c).$ 61

$v_L^{\vec{p}}$ 61

$c(i \mapsto k, k \mapsto i)$ 65

C 66

$\mathbb{H}_t$ 67

$Inj(n, m)$ 67

μ 67

$Contr(\alpha|c)$ 71

$u_L^{\vec{p}}$ 72

$\mathbb{B}$ 73

$\mathbb{H}_0$ 73

w_∞ 74

$B_k^{\vec{p}_n}$ 78

$E(c)$ 87

$\equiv^0 a_i a_j$ 89

$\equiv^1 a_i a_j$ 89

$QFSL^=$ 89

$EQ_p(L)$ 91

T_p 91

T_p^t 91

EQ_p^t 91

(SX=) 93

$u_{L=}^{\vec{p}}$ 94

$u_{=}^{\vec{p}}$ 94

$w(\varphi|\psi)$ 102

(PIR) 114

(JSP) 118

$g(\cdot, \cdot)$ 119

ID_p 120

$\langle n_1, n_2 \rangle$ 129

N 129

$\iota^{\#}$ 129

$\iota(H)$ 130

(PGC) 130

$vip_E(\beta)$ 131

$vip_E(\vec{m})$ 131

$E \leq_p F$ 131

ι_E 133

$N(\alpha, E, L)$ 134

$J(\mathbb{H})$ 137

(GPC) 137

$N\#(\vec{x}, \vec{y})$ 141