PROBABILISTIC RELATIONSHIPS, RELEVANCE AND IRRELEVANCE WITHIN THE FIELD OF UNCERTAIN REASONING

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

2007

Peter John Waterhouse
School of Mathematics

Contents

Abstract 5

Declaration 6

Copyright Statement 7

Acknowledgements 8

1 Introduction 9

1.1 Historical Context 9

1.2 The Urn Model 10

1.3 Concepts & Notation 11

1.3.1 Basics 11

1.3.2 Some Desirable Properties for w 14

1.4 Aims of This Thesis 29

2 Relationships for Probability Functions Which Satisfy Ax 30

2.1 Introduction & Motivation 30

2.2 Our Language 31

2.3 Probabilistic Relationships for Probability Functions Which satisfy Ex 32

2.4 State Descriptions That Differ by One Observation 33

2.4.1 The "Short Cut" 35

2.4.2 The "Long Road" 36

2.4.3 The Only Rule 47

3 Principles Involving Relevance **53**

3.1 Introduction & Motivation 53

3.2 Some Preliminaries 55

3.3 NPIR and SPIR Hold for the Carnap and Paris-Nix Continuums . . . 59

3.4 SPIR and All Probability Functions Satisfying Ax 64

3.4.1 Where Our Principles Hold 65

3.4.2 Where Our Principles Do Not Hold & Some Unresolved Cases 70

3.4.3 NPIR and the Unresolved Cases 84

3.4.4 A Quick Note on SPIR & Ex 88

3.4.5 Discussion 90

3.5 PIR and Ex 91

3.5.1 Historical Context 91

3.5.2 PIR does not imply Ex 92

3.5.3 The Most General Form of PIR 107

4 Principles Involving Irrelevance **112**

4.1 Introduction & Motivation 112

4.2 The Weak Irrelevance Principle 114

4.2.1 Our Language 114

4.2.2 A Single Constant 115

4.2.3 WIP and the Paris-Nix Continuum 118

4.2.4 Generating Further Probability Functions That Satisfy WIP . 123

4.3 Recoverable Probability Functions 131

5 Conclusions & Suggestions For Future Work **144**

5.1 The Only Rule 144

5.2 Principles Involving Relevance 146

5.2.1 NPIR & SPIR 146

5.2.2 PIR & Ex 146

5.3 Principles Involving Irrelevance 147

5.3.1 The Weak Irrelevance Principle 147

5.3.2 Recoverable Probability Functions 148
5.4 Overall 148

Bibliography **150**

A Mathematica Programme for Checking SPIR Counter-Example **154**

Word count 27570

The University of Manchester

Peter John Waterhouse
Doctor of Philosophy
Probabilistic Relationships, Relevance and Irrelevance within the Field of Uncertain Reasoning
December 28, 2007

In the 1950's Carnap kick-started the field of Uncertain Reasoning by publishing his formulation of a Continuum of inductive methods which satisfied several principles of symmetry, notably Johnson's Sufficientness Principle (which he had developed independently). Since then, relatively little work has been done in the area until a revival in the last decade, which this thesis aims to continue.

We start by reviewing the history of the area bringing the reader up to date, and providing a basis for the work to follow. The thesis can then be split into three subject areas.

The first of these will be an investigation into universal probabilistic relationships for functions which satisfy our main principle of symmetry, Atom exchangeability (Ax), the main result here being "The Only Rule", so named because we also prove it is the only rule of its type.

The second section will deal with principles involving relevance, and is largely based upon the Principle of Instantial Relevance (PIR). We introduce and discuss two new Principles, the Strong and Negative Principles of Instantial Relevance (SPIR & NPIR) both of which have a surprising sting in the tail for what we might regard to be common sense. Following this work there is some discussion regarding PIR and its relationship to Constant Exchangeability (Ex).

The third section deals with principles of irrelevance, notoriously a more difficult area to pin down. We study the Weak Irrelevance Principle (WIP), and which probability functions satisfy it. We then discuss Recoverable probability functions, the idea being that we can recover the function by conditioning on future events.

Finally we sum up our findings and comment on some future areas of research.

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 institute of learning.

Copyright Statement

Acknowledgements

I'd like to thank:

My supervisor, Jeff Paris, whose help and suggestions have been invaluable throughout the last three years.

My friends for keeping me entertained throughout, particularly Rosy, Adrian, Emma, Alsion, Linden, Mahinda, and Mike, who started this all off. Also included here are any of the various people who have been based in, or popped into, my office and anyone I have missed out by accident.

Dad, Mum, Anna & Grandma, of course.

And last, but definitely not least, Sarah, who has been fantastic throughout the last year and a half.

Chapter 1

Introduction

Our motivation in this chapter is to bring the reader up to date with regards to the areas of Inductive Logic that we will be concentrating on within this thesis, and to introduce the concepts and notation that we will be using throughout. To that end, we start with a historical summary of the subject and follow it with a longer section detailing the framework for what follows. Finally we sum up our aims for the thesis.

1.1 Historical Context

The question of 'What's Next?' has always held fascination for us, whether it be related to life or death matters, such as what to expect coming around that corner next when hunting, or more trivial pursuits, like who to bet upon in the upcoming derby.

Starting with the earliest of them, mathematicians have played their part in such matters, from the mythological numerology espoused by certain Greeks, to renaissance men such as Gerolamo Cardano, one of the men at the forefront of the work on cubic equations in the 16th century who predicted the date of his own death and then starved himself during the week prior to this date to ensure his claims veracity.

Of course mathematics has also played a more formal role in the attempts to address this question, notably within the fields of Mathematical Logic and Probability Theory.

It is here where our interest lies, more specifically within the field of Inductive Logic, where the question of 'What's Next' is modified to become:

"Assuming Nothing, what can we expect to see next based upon what we have already observed?"

The study of Inductive logic, based upon probabilistic theorems of Bayes and Cox, has progressed throughout the 20th century, initially through logicians such as William Johnson and Bruno de Finetti in the 1920's and 30's, then largely through the Carnapian (involving such logicians John Kemeny, Haim Gaifman and Rudolf Carnap himself) school of the 1950's and 60's.

Since Carnap more sporadic work has been done in the area before the recent upsurge, due in part to my own supervisor, which this thesis aims to continue.

1.2 The Urn Model

The urn model is a standard probabilistic tool which allows problems such as those we are about to address to be visualised.

Let there be 8 different[1] coloured balls, red, orange, yellow, etc.

Suppose we are picking, with replacement, these balls from a large urn and that we are completely ignorant to the proportions of the balls in this urn. Then

(1) Should we assign greater belief to picking 4 red and 1 green to picking 3 red

[1]We use 8 because it sits nicely with the forthcoming formalisation, where we will typically take our number of "options" to be some power of 2

and 2 green (in that order) in our first five picks?

(2) After we have picked 2 reds and a green should we give greater expectation to the 5th ball being red if the 4th ball is red rather than green?

If we replace our single urn by two and split the balls so that all the red, orange, yellow, green and blue balls reside in the first urn whilst the indigo, violet and black balls reside in the other then;

(3) Should the expectation of the next ball drawn from the first urn being red be unchanged, whatever we've drawn from the second urn?

General forms of these questions will be addressed in the forthcoming chapters.

We now formalise our notation and introduce some important concepts and definitions.

1.3 Concepts & Notation

We assume a very basic knowledge of predicate logic, such as can be found in [22]. Throughout what follows we will often provide an shortened name for each principle in brackets (for instance, Ex for Constant Exchangeability). These shorthand names will often be used throughout the thesis without allusion to their progenitors.

1.3.1 Basics

In this section we introduce the language structure we'll be using and the definition of a probability function.

For the most part we will be working with a language L for predicate logic with the constant symbols $a_1, a_2, \ldots$ and finitely many unary predicates, $P_1, P_2, \ldots, P_p$, but without function symbols or equality. The intended interpretation here is that these constants, which represent 'individuals', exhaust the universe. Let FL denote the formulae of L, SL the sentences of L and $QFSL$ the quantifier free sentences of L.

We are now in a position where we can define the concept of a probability function.

Definition 1.3.1. A map, $w : SL \to [0,1]$ is said to be a *probability function on* L if it satisfies that for all sentences $\theta, \phi, \exists x\, \psi(x) \in SL$:

(P1) If $\models \theta$ then $w(\theta) = 1$.
(P2) If $\models \neg(\theta \wedge \phi)$ then $w(\theta \vee \phi) = w(\theta) + w(\phi)$.
(P3)[2] $w\left(\exists x\, \psi(x)\right) = \lim_{m\to\infty} w\left(\bigvee_{i=1}^{m} \psi(a_i)\right)$.

Further, $w(\cdot \mid \cdot) : SL \times SL \to [0,1]$, is said to be the *conditional probability function on* L and is defined by

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

for $w(\phi) \neq 0$, equivalently

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

P1 and P2 are the predicate calculus versions of the traditional axioms for defining probability functions, and can be justified by, for example, the Dutch Book argument (developed by, amongst others, de Finetti and Ramsey in [8] and [28] and given as the second justification in chapter 3 of [26] (where two other justifications are also given)), which is based upon an agents "willingness to bet". P3 is considered a natural extension of P1 and P2 and is intended to capture the idea that the constants, a_i, constitute the whole of the universe (again see [26], chapter 11).

These Axioms were first formulated by Gaifman in [12], where he also showed that

[2]This axiom is also known as Gaifman's axiom.

any probability function defined on $QFSL$ satisfying P1 and P2 can be extended uniquely to a probability function on L satisfying P1, P2 and P3. This important theorem makes our work considerably easier, as we can limit our attention to concentrating on probability functions defined just upon the $QFSL$.

We next discuss two very useful concepts that help us structure our language in a clear way, Atoms (also called states) and State Descriptions.

Definition 1.3.2. Let $\alpha_1(x), \alpha_2(x), \ldots, \alpha_{2^p}(x)$ be an enumeration of the 2^p formulae of the form

$$\bigwedge_{i=1}^{p} P_i^{\epsilon}(x),$$

where $\epsilon \in \{0, 1\}$ and P_i^0, P_i^1 represent $\neg P_i$, P_i respectively. These $\alpha_i(x)$ are known as the *atoms* of the language L.

Definition 1.3.3. Given the atoms $\alpha_1(x), \alpha_2(x), \ldots, \alpha_{2^p}(x)$, a *state description*, is any sentence of the form

$$\phi(a_1, a_2, \ldots, a_n) = \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i).$$

These concepts are particularly useful for us with regards to the Disjunctive Normal Form Theorem (DNFT) stated below[3] in a way appropriate to this thesis.

Theorem 1.3.4. *Any sentence in $QFSL$ can be written as a disjunction of state descriptions of L.*

It follows that the probability function, w, can be entirely defined by the probability it awards to the state descriptions of the given language.

All elementary properties and conventions concerning probability functions will be assumed (see, for example [26]).

[3] For a proof of the DNFT see, for example, [27].

1.3.2 Some Desirable Properties for w

At first sight it seems as though there are too many probability functions for us to deal with, but we are only actually interested in those we might consider to be reasonable.

In order to formalise this notion we may imagine w to represent the beliefs assigned by some rational agent to the sentences of L in the absence of any further knowledge. It is essential to emphasize the "in the absence of any further knowledge" part of the last sentence, a condition which can easily be overlooked - especially when constructing seemingly counter-intuitive arguments (such as Goodman's Grue Paradox (see [1] for a definition or [15] for the original full statement), which is irrelevant to the work in this thesis due to this) as potential criticisms to this area of study.

Taking such a w allows us to specify some desirable properties for w (or we can think of them as principles which we may consider it irrational to ignore in the context of "the absence of any further knowledge"). A number of these, particularly those involving symmetry, will be fundamental to the work that follows in this thesis.

Our aim in this section, therefore, will be to introduce these desiderata, starting with those principles involving symmetry.

Symmetry Principles

We start with the most basic of symmetry principles - proposed independently by Carnap and Johnson in [2] and [20], respectively, Constant Exchangeability (Ex).

In simple terms, Ex asserts that the individuals identity doesn't matter in terms of our inductive analysis. That is to say, it doesn't matter who exhibits the property we're observing, merely that we have observed someone that exhibits that property.

As an example, consider the predicate P_1 =“Brown Hair”, then of the first three individuals we observe - Tom, Dick and Harry, two have brown hair. What probability should we assign to our fourth individual, Brian, having brown hair? Should this value depend which two of Tom, Dick and Harry had brown hair? The answer to the last question is No, particularly since we have no further knowledge (like, for instance, that Brian might be related to one or other of the previous subjects). We now define the concept mathematically.

Definition 1.3.5. For $\theta, \theta' \in SL$, if θ' is obtained by replacing the distinct constant symbols $a_{i1}, a_{i2}, \ldots, a_{in}$ in θ with the distinct constant symbols $a_{k1}, a_{k2}, \ldots, a_{kn}$ then *Constant Exchangeability* (Ex) asserts that $w(\theta) = w(\theta')$.

Ex allows us a simplification of notation for state descriptions. Since we no longer require (in order to determine the probability assigned to the state description) the information about which constant satisfies which atom we can just note how often each atom appears in the state description, and therefore unambiguously represent it via its vector representation, defined below.

Definition 1.3.6. Given a state description

$$\theta(a_1, a_2, \ldots, a_n) = \bigwedge_{j=1}^{n} \alpha_{h_j}(a_j),$$

its *Vector Representation* is given by $\vec{\theta} = \langle n_1, n_2, \ldots, n_{2^p} \rangle$ where $n_i = |\{\, j \mid h_j = i \,\}|$. (Note too that $\sum_{i=1}^{2^p} n_i = n$.)

In this case we may also express this state description as

$$\theta(a_1, a_2, \ldots, a_n) = \alpha_1^{n_1} \wedge \alpha_2^{n_2} \wedge \ldots \wedge \alpha_{2^p}^{n_{2^p}} = \bigwedge_{i=1}^{2^p} \alpha_i^{n_i}.$$

Finally, whenever we state $w(\vec{\theta})$ we mean $w(\theta)$.

The next symmetry principle we consider is Predicate Exchangeability (Px). Px states that it doesn’t matter what we’ve seen, what matters is how often we’ve seen it. So in our previous example, if we swap the predicate “brown hair” for “red hair” and then out of the first three individuals we've observed, two out of three of them had red hair, then the probability we assign to the 4th person having "red hair" should be equal to the probability of Brian having "brown hair" in the earlier example. Mathematically;

Definition 1.3.7. For $\theta, \theta' \in SL$, if θ' is obtained by replacing the (distinct) predicate symbols $P_{i1}, P_{i2}, \ldots, P_{ij}$ in θ with the (distinct) predicate symbols $P_{k1}, P_{k2}, \ldots, P_{kj}$, then *Predicate Exchangeability* (Px) asserts that $w(\theta) = w(\theta')$.

Similarly the principle of Strong Negation (SN) states that if we'd exchanged the predicate "red hair" by "not having red hair" in the above, and then observed two people without red hair, the probability that the fourth person would not have red hair should be, again, equal to the probability of the fourth person having red hair in our last example (also recall that we have no prior knowledge of the distribution of red heads in the population). Mathematically;

Definition 1.3.8. For $\theta, \theta' \in SL$, if P is any predicate symbol of L and θ' is obtained from θ by replacing each occurrence of P by $\neg P$ then *Strong Negation* (SN) asserts that $w(\theta) = w(\theta')$

From now on, unless otherwise stated[4], all probability functions we consider will be assumed to satisfy the standing assumptions of Ex, Px and SN.

Finally we come to our main exchangeability principle, Atom Exchangeability (Ax) which we will utilise throughout this thesis. Ax states that whole situations are mutable, that is to say for instance, that the probability of observing three constants with the properties that define α_1 is exactly the same as the probability we should assign to the probability of observing three constants with the properties that define α_2 or α_3 or etc. Mathematically;

[4]Notably, we will be dealing with probability functions that don't satisfy Ex towards the end of Chapter 3

Definition 1.3.9. Let σ be some permutation of $\{1, 2, \ldots 2^p\}$. Then *Atom Exchangeability* (Ax) states that

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

We now introduce the concept of a spectrum, first introduced in [23] over languages containing n-ary predicates, here we limit it to unary predicates.

Definition 1.3.10. Given a state description,

$$\theta(a_1, a_2, \ldots, a_n) = \bigwedge_{i=1}^{2^p} \alpha_i^{n_i},$$

its *Spectrum* is defined to be the multiset $\{n_1, n_2, \ldots, n_{2^p}\}$.

When a probability function satisfies Ax we can unambiguously represent a state description (in terms of the probabilistic value awarded to it) by its spectrum - an improvement on the simplification we earlier achieved through Ex.

Definition 1.3.11. We say that two state descriptions are *Exchangeable* if they have the same spectrum.

We now introduce a concept closely related to that of a spectrum and which is a direct extension of our vector representation of state descriptions idea.

Definition 1.3.12. Given a vector representation of a state description, $\vec{\theta} = \langle n_1, n_2, \ldots, n_{2^p}\rangle$, the *Ordered Vector Representation* of that state description is defined to be

$$\hat{\theta} = \langle n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(2^p)}\rangle,$$

where σ is a permutation of the $\{1, 2, \ldots 2^p\}$ such that $n_{\sigma(1)} \geq n_{\sigma(2)} \geq \ldots \geq n_{\sigma(2^p)}$, i.e. the ordered vector representation of a state description is the vector where the components of the vector representation of the same state description have been ordered in a non-increasing fashion.

Throughout the thesis we take $w(\hat{\theta})$ to mean $w(\theta)$.

Other Desiderata

In this section we mention some additional important principles, this time not based upon symmetry, but including one upon which Chapter 3 will largely be based upon. We start with the principle of Regularity.

Definition 1.3.13. *Regularity* (REG) states that for all state descriptions, θ we have that $w(\theta) > 0$. Equivalently $w(\psi) > 0$ for all non-contradictory $\psi \in QFSL$.

REG basically says that unless some situation is actually impossible (i.e. contradictory) then we should always give that event some possibility of occurring, even if we believe it to be very unlikely. In essence then, REG conveys the meaning of the famous saying "There's a first time for everything!".

The need for such a principle is due to the fundamental philosophical difference between a probability of 0, and that which is non-zero - namely that we shouldn't rule out as impossible any event which is actually possible. We will be mentioning this principle in relation to several of our standard probability functions satisfying Ax later in this chapter.

The following important principle was first suggested[5] by Carnap in [3].

Definition 1.3.14. For an atom $\alpha(x)$ and state description $\theta(a_1, a_2, \ldots, a_n)$ the *Principle of Instantial Relevance* (PIR) is the requirement that

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \theta(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \theta(a_1, a_2, \ldots, a_n)).$$

In essence, then, this principle states that

"If we have observed $\theta(a_1, a_2, \ldots a_n)$ then our belief in $\alpha(a_{n+2})$ holding should not be decreased if you also learn that $\alpha(a_{n+1})$ holds, i.e. learning that a_{n+1} satisfies α should (if anything) enhance your belief that a_{n+2} will too."

[5]For a historical summary regarding this principle, see section 3.5.1.

PIR can be shown to follow directly from Ex via de Finetti's theorem (to be explicitly stated shortly, also see for example [19]), thus confirming the apparent reasonableness (and therefore desirability) of the principle.

This principle will form the basis of Chapter 3 where we will discuss two related principles, the *Negative Principle of Instantial Relevance* (NPIR) and the *Strong Principle of Instantial Relevance* (SPIR). We will also be considering whether the reverse implication $PIR \to Ex$ holds, but we'll leave further discussion on these matters, including the definitions of the new principles, until we reach the chapter in question.

One related principle we will mention here is the Generalised Principle of Instantial Relevance, first mentioned in [17] and [18], and studied extensively in [23] and [25].

Definition 1.3.15. For $\theta(x), \phi(x), \psi(\vec{x}) \in QFSL$ and $\phi(a_{n+1}) \wedge \psi(a_1, \ldots a_n)$ consistent, if $\theta \models \phi$ then the *Generalised Principle of Instantial Relevance* (GPIR) requires that

$$w(\theta(a_{n+2}) \mid \phi(a_{n+1}) \wedge \psi(a_1, \ldots, a_n)) \geq w(\theta(a_{n+2}) \mid \psi(a_1, \ldots, a_n)).$$

In essence, then, GPIR extends the idea behind PIR, to say that the observation of an object satisfying a set of properties should not decrease the belief in further objects satisfying a superset of those properties. Further, it has been shown (in [18]) that GPIR is equivalent to the principle:

For $\theta(x), \phi(x), \psi(\vec{x}) \in QFSL$ and $\phi(a_{n+1}) \wedge \psi(a_1, \ldots, a_n)$ consistent, if $\phi \models \theta$ then

$$w(\theta(a_{n+2}) \mid \phi(a_{n+1}) \wedge \psi(a_1, \ldots, a_n)) \geq w(\theta(a_{n+2}) \mid \psi(a_1, \ldots, a_n)),$$

that is, exactly the same principle but with $\phi \models \theta$ instead of $\theta \models \phi$. This means that GPIR states also that the observation of an object satisfying a set of properties should not decrease the belief in further objects satisfying a subset of those properties.

So, if we observe a person who is tall with brown hair, say, then GPIR states that our expectation for the next person being tall (with indeterminate hair colour) should not be reduced but also that our expectation that the next person will be tall with brown hair and blue eyes should not be reduced. This would seem to satisfy our idea of common-sense, whilst undeniably being on shakier ground than PIR.

As it turns out, GPIR is an extremely strong condition - satisfied by only a narrow continuum of probability functions, the Paris-Nix Continuum, which we will discuss in the upcoming section on standard probability functions satisfying Ax.

Finally, for this section, we introduce the principle of Language Invariance, given here not in its most general form[6] but in a form appropriate to this thesis.

Definition 1.3.16. (The *Language Invariance Principle* (LIP)) Let w be a probability function defined on the language $L = \{P_1, P_2, \dots P_p\}$ and satisfying Ax. Then w is said to be Language Invariant if there are a family of probability functions $w_{\mathcal{L}}$ on $\mathcal{L}$ satisfying Ax for each finite language $\mathcal{L} = \{P_1, P_2, \dots, P_q\},\ q \in \mathbb{N}^+$ such that $w = w_L$ and if $\mathcal{L} \subseteq \mathcal{L}'$ then $w_{\mathcal{L}'} \restriction S\mathcal{L} = w_{\mathcal{L}}$.

In this case we say that these $w_{\mathcal{L}}$ form a Language Invariant family.

i.e. No matter how we set up our language, we ought to expect an equivalent result for equivalent sentences.

Naturally different people have different ideas as to what constitutes common sense, and as such many more “common-sense” principles have been proposed on top of the ones offered here, for example [17] contains definitions for several principles based upon irrelevance[7], such as the Strong, Conditional and Weak Irrelevance Principles

[6] A more general form can be found in [26], page 73.

[7] We will go into more detail on irrelevance principles in chapter 4 where the Weak Irrelevance Principle will get a thorough treatment.

(SIP, CIP and WIP, respectively), and the same thesis also defines the Reichenbach Axiom (RA) - which happens to contradict GPIR.

However, if we were to include all the principles that are out there, there wouldn't be any room for the original work to follow and we'd find ourselves perhaps a little too restricted in our choices of probability function (indeed, a review of all common sense principles could possibly fill a thesis of its own!). Therefore we have picked those which seem most relevant, and most sensible, but it is definitely worth bearing in mind that other principles exist and could be considered in any future work regarding what follows in this thesis.

The reader who is interested in what other principles (or indeed, what slightly different ways of thinking about how probability and logic are related) could be considered might consult, for instance, [17], [24], [31] or [21]. The latter contains an exposition of a very different way (non-Carnapian) of formalising Inductive Logic - called "New Fangled Inductive Logic" (NFIL) by Stephen Glaister in [14], his overview of the subject to date.

There is also one principle we have not yet introduced (that we intend to in this introduction), namely Johnson's Sufficientness Principle (JSP). This will be defined and explained in the forthcoming section on some standard probability functions that satisfy Ax[8].

de Finetti's Theorem

The importance we assign to the usefulness of de Finetti's Theorem, taken from [9] (also see [10], [11]), can be garnered from the fact that we believe it deserves to be stated in its own section.

The version we state here was proven in [7] and relies solely on the fact that our

[8]Where it will also become clear why we have left it until then!

probability function satisfies the relatively undemanding Ex.

Theorem 1.3.17. *(de Finetti's Representation Theorem) Let*

$$\mathbb{D}_{2^p} = \left\{ \langle x_1, x_2, \ldots, x_{2^p} \rangle \in \mathbb{R}^{2^p} \mid x_i \geq 0, \sum_i x_i = 1 \right\}.$$

If w satisfies Ex, then there is a normalised countably additive measure μ on the Borel subsets of $\mathbb{D}_{2^p}$ such that

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \int_{\mathbb{D}_{2^p}} \prod_{r=1}^{2^p} x_r^{n_r} d\mu(\vec{x}),$$

where $n_r = |\{k \mid h_k = r\}|$ for $r = 1, 2, \ldots, 2^p$.

i.e. Any probability function, w, which satisfies Ex is associated with a unique prior distribution function, μ, on $\mathbb{D}_{2^p}$.

This powerful and elegant result allows us to solve several problems which would otherwise be far more difficult. Indeed - it is used to prove that PIR follows directly from Ex and we will be using it ourselves in this thesis for a similar purpose.

Note that whenever we use integrals in the sequel it will generally be in this context. Therefore, we will omit the $\mathbb{D}_{2^p}$ unless we intend to integrate over some other set.

Some Standard Probability Functions That Satisfy Ax

We now move on to discussing some of the standard probability functions that satisfy Ax. Many of these will be mentioned in the work that follows, and the two continuums will be utilised extensively.

We're going to initially introduce Carnap's $\mathfrak{m}_o$ (see [4]), quite frankly a dreadfully degenerative probability function, but also one which appears at the limits for both of our continuum's, and therefore a function that we need to mention. But first, a new concept;

Definition 1.3.18. A *Homogenous State Description* is a state description where every constant is satisfied by the same atom. i.e.

$$\theta(a_1, \ldots, a_n) = \bigwedge_{i=1}^{n} \alpha_I(a_i)$$

where $1 \leq I \leq 2^p$ is some constant, defines a homogenous state description. A *Non-Homogenous State Description* is a state description of any other type (i.e. where more than one atom has been observed).

Definition 1.3.19. For the state description $\theta \in SL$, *Carnap's* $\mathfrak{m}_o$ is defined to be the probability function where

$$w(\theta) = \begin{cases} \frac{1}{2^p} & \text{for homogenous } \theta, \\ 0 & \text{otherwise.} \end{cases}$$

i.e. as soon as you've made your first observation, you will then expect your second, third, ..., $n'th$ observation to be the same. Obviously this probability function is extremely undesirable for our purposes, and we'll only be mentioning it during the thesis in regards to exceptional cases.

We now move on to a slightly more sensible probability function, the Independent Solution, w_I. You might imagine that since we've stipulated (through our symmetry principles) that every atom needs to be treated in exactly the same way in terms of the probability we assign, that a natural way to assign belief is to let every state description have the same value for each level of n (the number of constants), and indeed *any* probability function that satisfies Ax must award each atom (i.e. a state description over one constant) the same probability, namely $\frac{1}{2^p}$. The definition of the probability function now follows.

Definition 1.3.20. For any state description $\theta(a_1, a_2, \ldots, a_n) \in SL$, *The Independent Solution,* w_I is defined to be the probability function where

$$w(\theta(a_1, a_2, \ldots, a_n)) = \left(\frac{1}{2^p}\right)^n.$$

The Independent Solution is ideal if all we are doing is tossing a fair coin. It is true that the independent solution is the most symmetric probability function you can come up with as it treats every situation without prejudice, but this is exactly our problem - the whole point of inductive learning is that you are learning! Under w_I if you had tossed a coin 1000 times and it came up heads 1000 times, you'd still have to give the probability of the next toss bringing tails a value of $\frac{1}{2}$ - This is clearly unacceptable as a rational agent would by now consider that the coin is loaded to land on heads, and would give a higher probability to such an outcome.

We therefore reject w_I because it fails to be inductive and as such we would like to introduce some inductive learning into our probability function. Towards that end, Johnson in [20] and Carnap in [3], [4] (independently of each other) came up with what we now know as Johnson's Sufficientness Principle (JSP):

Definition 1.3.21. For an atom $\alpha_i \in SL$ and a state description $\theta(a_1, a_2, \ldots, a_n) = \bigwedge_{j=1}^n \alpha_{h_j}(a_j) \in SL$, *Johnson's Sufficientness Principle* (JSP) states that the probability

$$w\left(\alpha_i(a_{n+1}) \mid \bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right),$$

is a function only of n and $|\{\, j \mid h_j = i \,\}|$, the number of times α_i appears in θ.

It turns out that JSP gives rise to a famous continuum of probability functions, but more on that shortly - first we are going to consider Carnap's $\mathfrak{c}_0$ (see [4]), more commonly known as the "Straight Rule", which would appear to be the most straightforward exponent of JSP.

Definition 1.3.22. For an atom $\alpha_i \in SL$ and a state description $\theta(a_1, a_2, \ldots, a_n) = \bigwedge_{j=1}^n \alpha_{h_j}(a_j) \in SL$, *Carnap's* $\mathfrak{c}_0$ states that

$$w\left(\alpha_i(a_{n+1}) \mid \bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) = \frac{m}{n},$$

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

i.e. The straight rule gives the probability of the next observation being α_i exactly equal to the proportion of times we have seen that atom so far. This has several problems. The most obvious one is that our initial probabilities are undefined - when nothing has happened we cannot assign a probability to any future event, indeed the Straight Rule fails to adhere to the laws of probability and doesn't actually define a probability function.

Another problem is that the Straight Rule fails to satisfy REG, whenever we have not observed an event we are forced to give the probability of its future observation to be 0 - i.e. we are forced to conclude that such an event is *impossible*. This may be vaguely feasible (though still undesirable) if we'd observed the 1000 heads in a row (from our earlier coin tossing experiment) but the real killer here is that it behaves in the same way as $\mathfrak{m}_\circ$ (indeed, essentially $\mathfrak{c}_\circ$ is equivalent to $\mathfrak{m}_\circ$) in that once we've made a single observation, we are forced to conclude that the next observation will be the same (since that now has probability one, and every other atom has probability zero), clearly the straight rule is undesirable.

However, the idea of the proportion of past observations affecting our belief in future observations is crucial to inductive learning - and as such JSP does give rise to a family of functions which are more sensible and have several nice properties. This family is known as Carnap's Continuum, and was introduced in [4].

Definition 1.3.23. *Carnap's λ-Continuum* The only possible probability functions that satisfy Ex, Ax and JSP are $\mathfrak{m}_\circ$, w_I and the solutions w_λ determined by

$$w_\lambda\left(\bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) = \frac{\prod_{r=1}^{2^p}\prod_{k=0}^{n_i-1}(k+\frac{\lambda}{2^p})}{\prod_{k=0}^{n-1}(k+\lambda)},$$

where $n_i = |\{\, j \mid h_j = i \,\}|$ for $i = 1, 2, \ldots, 2^p$, and $\lambda \in (0, \infty)$.

The formulation is slightly easier when we consider the continuum in relation to conditional probability functions, where it becomes

$$w_\lambda\left(\alpha_i(a_{n+1}) \mid \bigwedge_{j=1}^{n} \alpha_{h_j}(a_j)\right) = \frac{m+\frac{\lambda}{2^p}}{n+\lambda},$$

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

In Carnap's continuum the parameter, λ, can be seen as striking a balance between the empirical factor and the logical factor that goes into deciding what belief we'd give towards any particular state description. In some respects λ can be likened to the number of previous experiments we have run where all the atoms have been observed an equal number of times. Indeed, if we allow λ to tend towards its limits we get the Straight Rule answer that once we've observed one atom we give all the probability for future observations to observing the same atom again, namely the continuum tends towards $\mathfrak{m}_o$ at its lower limit. We also note that if we allow λ to tend towards its upper limit of ∞ we get w_I where whatever we observe, we expect each of the possible atoms to occur next with equal probability. The continuum does provide some more sensible probability functions between the limits though.

The Continuum has several attractive properties, for instance it satisfies all our principles of symmetry, and REG except at its lower limit. There have been some arguments made against it, however, such as its inability to cope with universal generalisations (see, for instance, [16] and [30]). An interesting critique of the continuum, where it is compared to the Paris-Nix Continuum, which we will be introducing shortly) can be found comprising Chapter 4 of [23].

As mentioned earlier, GPIR leads us to a narrow range of probability functions, based upon one parameter, introduced in [23] and [25] which also satisfies several attractive properties - The Paris-Nix Continuum, which we will now define.

Definition 1.3.24. Let $\langle n_1, n_2, \ldots n_{2^p} \rangle$ be the vector representation of the state description, $\theta \in SL$. Then the probability given to θ by the *Paris-Nix Continuum* is defined to be

$$w^\delta(\vec{\theta}) = \frac{1}{2^p} \sum_{i=1}^{2^p} (\gamma + \delta)^{n_i} \gamma^{n-n_i}$$

where $2^p \gamma = 1 - \delta$ and $\delta \in [0, 1]$.

An alternative statement of this, when[9] $\gamma \neq 0$, which matches that originally given in [23] is; With $\vec{\theta}$ given as above, the probability awarded by the Paris-Nix Continuum is defined to be

$$w^{\delta}(\vec{\theta}) = \frac{\gamma^n}{2^p} \sum_{i=1}^{2^p} \left(1 + \frac{\delta}{\gamma}\right)^{n_i}$$

where $2^p\gamma = 1 - \delta$ and $\delta \in [0, 1)$. We will be using both definitions as appropriate throughout the thesis.

Again the limits of this continuum are undesirable - this time w^0 defines the Independent Solution and w^1 Carnap's $\mathfrak{m}_{\mathfrak{o}}$, but again when we pick a value of δ within the limits the result is a probability function which satisfies all the symmetry principles, REG, and has several other nice properties too. Several of these are discussed throughout Chapters 3 and 4 in [23].

We now look at a method of generating probability functions satisfying Ax.

Definition 1.3.25. For de Finetti measure μ defined over the Borel Subsets, B, of $\mathbb{D}_{2^p}$ let the corresponding Symmetrized Measure (SM) probability function be defined by the de Finetti measure $\overline{\mu}$, where

$$\overline{\mu} = \frac{1}{(2^p)!} \sum_{\tau} \mu_{\tau},$$

where τ ranges over the permutations of $\{1, 2, \dots 2^p\}$ and $\mu_\tau(B) = \mu(B_\tau)$ where

$$\begin{aligned} B_\tau &= \{\langle x_1, x_2, \dots, x_{2^p}\rangle \mid \langle x_{\tau^{-1}(1)}, x_{\tau^{-1}(2)}, \dots, x_{\tau^{-1}(2^p)}\rangle \in B\} \\ &= \{\langle x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(2^p)}\rangle \mid \langle x_1, x_2, \dots, x_{2^p}\rangle \in B\}. \end{aligned}$$

i.e. We are averaging over all the permutations of the measure, μ, to produce a new measure $\overline{\mu}$. Since the defining conditions for a probability function are closed under convex combination (as they are linear) this defines another probability function.

Theorem 1.3.26. *SM-Functions satisfy Ax.*

[9]When $\gamma = 0$ (i.e. $\delta = 1$) we can either use the first definition or our knowledge that w^1 is equivalent to $\mathfrak{m}_{\mathfrak{o}}$.

Proof. We need to show that any two exchangeable state descriptions are given the same probability by any SM probability function.

Let μ, τ and $\overline{\mu}$ be as defined above with w the (SM) probability function defined by the measure $\overline{\mu}$. Let $\theta, \phi \in SL$ be two state descriptions such that

$$\begin{aligned}\vec{\theta} &= \langle n_1, n_2, \ldots, n_{2^p} \rangle, \\ \vec{\phi} &= \langle n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(2^p)} \rangle,\end{aligned}$$

for some arbitrary permutation σ of $\{1, 2, \ldots, 2^p\}$. Then

$$\begin{aligned} w(\theta) &= \int x_1^{n_1} x_2^{n_2} \ldots x_{2^p}^{n_{2^p}} d\overline{\mu} \\ &= \frac{1}{(2^p)!} \sum_{\tau} \int x_{\tau(1)}^{n_1} x_{\tau(2)}^{n_2} \ldots x_{\tau(2^p)}^{n_{2^p}} d\mu \\ &= \frac{1}{(2^p)!} \sum_{\tau} \int x_{\tau\sigma^{-1}(1)}^{n_1} x_{\tau\sigma^{-1}(2)}^{n_2} \ldots x_{\tau\sigma^{-1}(2^p)}^{n_{2^p}} d\mu, \end{aligned}$$

since the $\tau\sigma^{-1}$ also run through all permutations of $\{1, 2, \ldots, 2^p\}$,

$$\begin{aligned} &= \frac{1}{(2^p)!} \sum_{\tau} \int x_{\tau(1)}^{n_{\sigma(1)}} x_{\tau(2)}^{n_{\sigma(2)}} \ldots x_{\tau(2^p)}^{n_{\sigma(2^p)}} d\mu \\ &= \int x_1^{n_{\sigma(1)}} x_2^{n_{\sigma(2)}} \ldots x_{2^p}^{n_{\sigma(2^p)}} d\overline{\mu} \\ &= w(\phi). \end{aligned}$$

Since we picked our permutation, σ, arbitrarily we have that this holds for all such permutations. Therefore whenever θ and ϕ are exchangeable they are given the same probability function by any SM probability function, as required. $\square$

In this way, then, we can create a probability function, symmetric under Ax, based upon a measure which needn't originally have satisfied Ax.

SM probability functions will be cropping up throughout the thesis, particularly as a means to produce probability functions which form counter-examples to various ideas.

1.4 Aims of This Thesis

This section outlines what we are trying to achieve throughout this thesis.

We aim to uncover some of the underlying structure behind symmetric probability functions, by studying the way in which probabilities of some state descriptions are related to each other. This will be done in the chapter 2, which follows this section.

We aim to provide insight into some new principles, such as SPIR & NPIR in chapter 3 and the concept of Recoverable probability functions in chapter 4. We also aim to provide insight into one principle which have been introduced before, but has not been studied in depth, WIP in chapter 4.

We aim to help answer unsolved questions within the area, in particular on whether the principle PIR implies Ex, the converse having been known since the 1950's, towards the end of chapter 3.

Finally we aim to suggest some ideas, based upon the work in this thesis, for future research. This will be done through chapter 5 at the end of the thesis.

Our overall aim is to provide a thesis which addresses several areas of Uncertain Reasoning, pushing each a little way forward in order to provide a significant overall contribution.

Chapter 2

Relationships for Probability Functions Which Satisfy Ax

2.1 Introduction & Motivation

Our motivation for this chapter will be to study the probabilistic relationships between state descriptions in order to attempt to discover some of the underlying structure beneath symmetric probability functions.

Specifically, our motivating question can be summed up by

"Given two arbitrary state descriptions, what conditions necessitate that one of them is always awarded a larger probability given that our probability function is symmetric?"

We start by looking at the case where our probability functions are restricted to satisfying Ex, a problem that turns out to be relatively trivial, and then progress to looking at the case where our probability functions must satisfy Ax, where the problem is decidedly non-trivial.

For this latter case we start by considering what the situation is between two state descriptions which differ by just one observation (exactly what we mean by this will be defined at the time), and it turns out that we can say something about which of the two will have the greater probability.

We offer two proofs of this result. The first is straightforward and relies upon de Finetti's theorem - we call this method the "Short cut". The second, which does not use de Finetti's theorem, is much trickier - we call this method the "Long road".

The immediate question here is "Why is it necessary to give two methods for the same proof?" our full reasoning will be given at the time, but one of the main reasons is that the "Long road" gives a greater amount of insight as to why the result holds.

Once our proofs to the result involving state descriptions which differ by only one observation are in place we are able to define two new concepts, the m-step and the m-route, which lead directly to the main result of the chapter "The Only Rule".

This result gives the condition required to be able to say that one state description will always have a greater than or equal probability to another whenever our probability function satisfies Ax. Thus addressing the motivating question for this chapter.

2.2 Our Language

Throughout this chapter let L be the language that contains the predicates $P_1(x), P_2(x), \ldots, P_p(x)$. Let the $\alpha_i(x)$ list the atoms of L for $1 \leq i \leq 2^p$.

We will be regularly be using two state descriptions, θ, $\phi \in SL$ which have vector representations given by

$$\vec{\theta} = \langle n_1, n_2, \ldots, n_{2^p} \rangle,$$

$$\vec{\phi} = \langle m_1, m_2, \ldots, m_{2^p} \rangle.$$

We will need to refer to the number of observations in each of these state descriptions, so let $N = \sum_i n_i$ and $M = \sum_i m_i$ in what follows.

2.3 Probabilistic Relationships for Probability Functions Which satisfy Ex

We start by briefly considering the following question:

"Given state descriptions, $\theta, \phi \in SL$, what conditions on them are sufficient to give $w(\theta) \geq w(\phi)$ whenever w is a probability function satisfying Ex?"

Theorem 2.3.1. *Given θ, ϕ as defined in section 2.2, a necessary and sufficient condition such that we have*

$$w(\theta) \geq w(\phi)$$

whenever w satisfies Ex is that $m_i \geq n_i$ for all $1 \leq i \leq 2^p$.

Proof. We need to consider both directions. First we will show that if we do not have that $m_i \geq n_i$ for all $1 \leq i \leq 2^p$ that there exists a probability function satisfying Ex such that $w(\phi) \geq w(\theta)$.

Assume we have that $m_i < n_i$ for some $1 \leq i \leq 2^p$. Let w be the probability function defined by the de Finetti measure μ which all the measure on the point

$$\langle \tau, \dots, \tau, \epsilon, \tau, \dots, \tau \rangle \in \mathbb{D}_{2^p},$$

where $\tau = \frac{1-\epsilon}{2^p-1}$, $\tau \gg \epsilon$ and the ϵ appears as the i'th coordinate. This probability function trivially satisfies Ex since it is defined by a de Finetti measure. Here we have that

$$w(\theta) = \tau^{N-n_i}\epsilon^{n_i},$$

$$w(\phi) = \tau^{M-m_i}\epsilon^{m_i}.$$

Since we can take ϵ as small as we like, we will always be able to find a probability function, w, such that

$$w(\theta) < w(\phi),$$

as required for this direction.

For the other direction, assume that we have $m_i \geq n_i$. By de Finetti we have that

$$w(\theta) = \int x_1^{n_1} x_2^{n_2} \dots x_{2^p}^{n_{2^p}} \, d\mu,$$
$$w(\phi) = \int x_1^{m_1} x_2^{m_2} \dots x_{2^p}^{m_{2^p}} \, d\mu.$$

Trivially we have that

$$\int x_1^{n_1} x_2^{n_2} \dots x_i^{n_i} \dots x_{2^p}^{n_{2^p}} \, d\mu \geq \int x_1^{n_1} x_2^{n_2} \dots x_i^{n_i+1} \dots x_{2^p}^{n_{2^p}} \, d\mu,$$

for any $1 \leq i \leq 2^p$, and therefore by extending this, we must have that

$$\int x_1^{n_1} x_2^{n_2} \dots x_i^{n_i} \dots x_{2^p}^{n_{2^p}} \, d\mu \geq \int x_1^{m_1} x_2^{m_2} \dots x_{2^p}^{m_{2^p}} \, d\mu,$$

i.e.

$$w(\theta) \geq w(\phi),$$

as required. Therefore both directions are shown and the theorem is proven. □

Particularly note that as a consequence of this theorem, if $N = M$ then the only time we can say anything about the relative probabilities of the two state descriptions for all probability functions satisfying Ex is the rather trivial case of when the state descriptions have identical vector representations, when they will be awarded identical probabilities.

2.4 State Descriptions That Differ by One Observation

We now move on to the more interesting case where w must satisfy Ax.

Let $\theta \in SL$ be defined as in section 2.2 and let $\phi \in SL$ be exchangeable with θ except that we replace one observation with another which has been observed less often. i.e. where θ has the ordered vector representation

$$\hat{\theta} = \langle n_1, n_2, \dots, n_{2^p} \rangle,$$

where the permutation we use allows that j is minimal and i maximal amongst the components with the same magnitude, ϕ has the ordered vector representation

$$\hat{\phi} = \langle n_1, n_2, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_{2^p} \rangle,$$

where $1 \leq i \leq 2^p - 1,\ 2 \leq j \leq 2^p,\ n_i > n_j + 1$.

Note that if we take $n_i = n_j + 1$ then both state descriptions would have the same ordered vector representation (i.e. n_i and n_j would swap values), and therefore would have an equal probability. It will therefore only be necessary to consider when this is not the case. The aim of the next two sections will be to provide proofs (one which relies on de Finetti's Theorem, and the other which does not) for the following theorem.

Theorem 2.4.1. *If we have θ, ϕ as defined above then for any probability function satisfying Ax we have that*

$$w(\theta) \geq w(\phi). \tag{2.1}$$

The reader may question why it is necessary to provide two proofs for the same Theorem, or more exactly, why the longer more difficult proof is necessary. Our reasoning is two-fold:

(1) Whilst the work in this thesis is purely concerned with languages containing only unary predicates, we hope that the work will in some part be transferrable to languages where an extended version of de Finetti's theorem may not be available[1].

(2) Whilst the use of de Finetti's theorem makes the problem much easier, it does so by taking out the intermediary steps which are present in the second method. These intermediary steps may well allow a greater level of mathematical insight as to why the result holds.

[1]It has now been brought to our attention that a yet to be published extended version of de Finetti's theorem now exists for higher arity languages. However, as yet this has not yet yielded a higher arity version of the Only Rule.

Once this result is shown to hold we will be in a position to state and prove The Only Rule.

2.4.1 The "Short Cut"

Here we can go straight into the proof of Theorem 2.4.1

Proof. Let ψ be the state description defined by

$$\theta = \alpha_i^{n_i} \wedge \alpha_j^{n_j} \wedge \psi,$$

i.e. ψ is the state description equivalent to all the observations in θ which do not involve α_i or α_j. Then we have that (2.1) is equivalent to

$$w(\alpha_i^{n_i} \wedge \alpha_j^{n_j} \wedge \psi) \geq w(\alpha_i^{n_i-1} \wedge \alpha_j^{n_j+1} \wedge \psi). \tag{2.2}$$

Applying de Finetti's theorem to this gives that, for some measure μ, we have that

$$\int x_i^{n_i} x_j^{n_j} f(\psi) d\mu \geq \int x_i^{n_i-1} x_j^{n_j+1} f(\psi)\, d\mu, \tag{2.3}$$

where

$$f(\psi) = \prod_{\substack{1 \leq k \leq 2^p \\ k \neq i,j}} x_k^{n_k}.$$

Now we have that (2.3) is equivalent to

$$\begin{aligned} \int x_i^{n_i} x_j^{n_j} f(\psi) d\mu - \int x_i^{n_i-1} x_j^{n_j+1} f(\psi)\, d\mu &\geq 0 \\ \iff \int \left(x_i^{n_i} x_j^{n_j} - x_i^{n_i-1} x_j^{n_j+1} \right) f(\psi)\, d\mu &\geq 0. \end{aligned} \tag{2.4}$$

If by Ax we swap α_i and α_j in (2.2) then the equivalent derivation leads us to

$$\int \left(x_j^{n_i} x_i^{n_j} - x_j^{n_i-1} x_i^{n_j+1} \right) f(\psi)\, d\mu \geq 0. \tag{2.5}$$

We now add (2.4) and (2.5) together to give

$$\begin{aligned} \int \left(x_i^{n_i} x_j^{n_j} + x_j^{n_i} x_i^{n_j} - x_i^{n_i-1} x_j^{n_j+1} - x_j^{n_i-1} x_i^{n_j+1} \right) f(\psi)\, d\mu &\geq 0 \\ \iff \int \left(x_i^{n_i-1} x_j^{n_j} (x_i - x_j) + x_i^{n_j} x_j^{n_i-1} (x_j - x_i) \right) f(\psi)\, d\mu &\geq 0 \\ \iff \int \left((x_i - x_j)(x_i^{n_i-1} x_j^{n_j} - x_i^{n_j} x_j^{n_i-1}) \right) f(\psi)\, d\mu &\geq 0 \\ \iff \int \left(x_i^{n_j} x_j^{n_j} (x_i - x_j)(x_i^{n_i-n_j-1} - x_j^{n_1-n_j-1}) \right) f(\psi)\, d\mu &\geq 0. \end{aligned} \tag{2.6}$$

To show that (2.6) holds it is clearly enough to show that (since $f(\psi)$ will always be non-negative)

$$x_i^{n_j} x_j^{n_j}(x_i - x_j)(x_i^{n_i-n_j-1} - x_j^{n_1-n_j-1}) \geq 0,$$

for all $0 \leq x_i, x_j \leq 1$. Looking at each multiplicand individually we have that

$$x_i^{n_j} x_j^{n_j} \geq 0, \tag{2.7}$$

and either

$$x_i - x_j \geq 0 \text{ and } x_i^{n_i-n_j-1} - x_j^{n_1-n_j-1} \geq 0,$$

or

$$x_i - x_j \leq 0 \text{ and } x_i^{n_i-n_j-1} - x_j^{n_1-n_j-1} \leq 0,$$

since $n_i > n_j$. Both of these possibilities give that

$$(x_i - x_j)(x_i^{n_i-n_j-1} - x_j^{n_1-n_j-1}) \geq 0,$$

and adding this fact to (2.7) gives that

$$x_i^{n_j} x_j^{n_j}(x_i - x_j)(x_i^{n_i-n_j-1} - x_j^{n_1-n_j-1}) \geq 0,$$

as required. $\square$

2.4.2 The "Long Road"

Here we have to work through a couple of lemmas before we will be in a position to prove Theorem 2.4.1. We first need a new definition;

Definition 2.4.2. Let ψ_1, $\psi_2 \in SL$ be state descriptions, then the *Routes* function, denoted

$$F(\vec{\psi_1}, \vec{\psi_2}),$$

is equal to the number of ways that we can extend ψ_1 such that it will be exchangeable with ψ_2.

As before, and throughout this section, we take θ such that

$$\hat{\theta} = \langle n_1, n_2, \ldots, n_i, \ldots, n_j, \ldots, n_{2^p} \rangle,$$

where we take the permutation which gives j maximal and i minimal amongst components of the same magnitude, and ϕ so that

$$\hat{\phi} = \langle n_1, n_2, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_{2^p} \rangle,$$

where $1 \leq i \leq 2^p - 1,\ 2 \leq j \leq 2^p,\ n_i > n_j + 1$, i.e. θ and ϕ differ by one observation (and are not exchangeable).

Lemma 2.4.3. *Let θ, ϕ be defined as above. Let*

$$B_s^K = \left\{ \vec{k} \mid \left(\sum_i k_i = K \right) \wedge (\{ i \mid 0 < k_i \leq s \} = \emptyset) \wedge (k_1 \geq k_2 \geq \ldots \geq k_{2^p}) \right\}$$

for $s \gg \max n_i$. Then there is a positive real number b_s, with $\lim_{s \to \infty} b_s = 0$ and, for any state description, ψ, such that $\hat{\psi} \in B_s^K$, we have;

$$F(\vec{\theta}, \vec{\psi}) > (1 - b_s) F(\vec{\phi}, \vec{\psi}). \tag{2.8}$$

unless $\vec{\psi}$ has more zero coordinates than $\vec{\theta}$, in which case both sides of (2.8) are equal to zero.

Proof. We first consider the case where $k_i \neq 0$ for all i. Let τ range over all permutations of $\langle 1, \ldots, 2^p \rangle$. Let $F'(\vec{\theta}, \hat{\psi})$ be defined as the routes function where the individual components are fixed (i.e. n_1 must represent the same state description as k_1), then we have that

$$\begin{aligned} & F\left(\langle n_1, \ldots, n_{2^p} \rangle, \langle k_1, \ldots, k_{2^p} \rangle\right) \\ = & \sum_\tau F'\left(\langle n_{\tau(1)}, \ldots, n_{\tau(2^p)} \rangle, \langle k_1, \ldots, k_{2^p} \rangle\right). \end{aligned}$$

Let τ_{ab} range over all permutations of $\langle 1, \ldots, i, \ldots, j, \ldots, 2^p \rangle$ where $\tau_{ab}(a) = i$ and $\tau_{ab}(b) = j$, and i and j are the coordinates which relate to the "different" observation in θ and ϕ. Then we have that

$$\begin{aligned} & \sum_\tau F'\left(\langle n_{\tau(1)}, \ldots, n_{\tau(2^p)} \rangle, \langle k_1, \ldots, k_{2^p} \rangle\right) \\ = & \sum_{a<b} \sum_{\tau_{ab}} F'\left(\langle n_{\tau_{ab}(1)}, \ldots, n_{\tau_{ab}(2^p)} \rangle, \langle k_1, \ldots, k_{2^p} \rangle\right) \\ & \qquad + F'\left(\langle n_{\overline{\tau}_{ab}(1)}, \ldots, n_{\overline{\tau}_{ab}(2^p)} \rangle, \langle k_1, \ldots, k_{2^p} \rangle\right), \end{aligned}$$

where within each summand $\overline{\tau}_{ab}$ is τ_{ab} composed with the transposition that maps a to b and b to a. We are going to show that

$$\begin{aligned}
&\sum_{a<b}\sum_{\tau_{ab}} F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_i,\ldots,n_j,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&\qquad +F'\left(\langle n_{\overline{\tau}_{ab}(1)},\ldots,n_j,\ldots,n_i,\ldots,n_{\overline{\tau}_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
> &\ (1-b_s)\sum_{a<b}\sum_{\tau_{ab}} F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_i-1,\ldots,n_j+1,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&\qquad +F'\left(\langle n_{\overline{\tau}_{ab}(1)},\ldots,n_j+1,\ldots,n_i-1,\ldots,n_{\overline{\tau}_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right),
\end{aligned} \tag{2.9}$$

note that this is equivalent to (2.8). We arbitrarily pick one pair of values $\{a,b\}$, with $1\le a<2^p,\ 1<b\le 2^p,\ a<b$, and aim to show:

$$\begin{aligned}
&F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_i,\ldots,n_j,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&+F'\left(\langle n_{\overline{\tau}_{ab}(1)},\ldots,n_j,\ldots,n_i,\ldots,n_{\overline{\tau}_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
> &\ (1-b_s)\left(F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_i-1,\ldots,n_j+1,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\right.\\
&\qquad \left.+F'\left(\langle n_{\overline{\tau}_{ab}(1)},\ldots,n_j+1,\ldots,n_i-1,\ldots,n_{\overline{\tau}_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\right).
\end{aligned} \tag{2.10}$$

We have that

$$\begin{aligned}
&F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_i,\ldots,n_j,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&\quad=\frac{(K-N)!}{(k_1-n_{\tau_{ab}(1)})!\ldots(k_a-n_i)!\ldots(k_b-n_j)!\ldots(k_{2^p}-n_{\tau_{ab}(2^p)})!},
\end{aligned}$$

$$\begin{aligned}
&F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_j,\ldots,n_i,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&\quad=\frac{(K-N)!}{(k_1-n_{\tau_{ab}(1)})!\ldots(k_a-n_j)!\ldots(k_b-n_i)!\ldots(k_{2^p}-n_{\tau_{ab}(2^p)})!},
\end{aligned}$$

$$\begin{aligned}
&F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_a-1,\ldots,n_b+1,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&\quad=\frac{(K-N)!}{(k_1-n_{\tau_{ab}(1)})!\ldots(k_a-(n_i-1))!\ldots(k_b-(n_j+1))!\ldots(k_{2^p}-n_{\tau_{ab}(2^p)})!},
\end{aligned}$$

$$\begin{aligned}
&F'\left(\langle n_{\tau_{ab}(1)},\ldots,n_b-1,\ldots,n_a+1,\ldots,n_{\tau_{ab}(2^p)}\rangle,\langle k_1,\ldots,k_{2^p}\rangle\right)\\
&\quad=\frac{(K-N)!}{(k_1-n_{\tau_{ab}(1)})!\ldots(k_a-(n_j+1))!\ldots(k_b-(n_i-1))!\ldots(k_{2^p}-n_{\tau_{ab}(2^p)})!}.
\end{aligned}$$

Each of these has a common factor of

$$\frac{(K-N)!}{\prod_{d\neq a,b}(k_d - n_{\tau_{ab}(d)})!}\ ,$$

so when we plug them into (2.10) we get

$$\begin{aligned}
& \frac{1}{(k_a-n_i)!(k_b-n_j)!} + \frac{1}{(k_a-n_j)!(k_b-n_i)!} \\
> \ & (1-b_s)\left(\frac{1}{(k_a-(n_i-1))!(k_b-(n_j+1))!} + \frac{1}{(k_a-(n_j+1))!(k_b-(n_i-1))!}\right)
\end{aligned}$$

$$\begin{aligned}
\Longleftrightarrow \quad & \frac{(k_a-n_j)!(k_b-n_i)! + (k_a-n_i)!(k_b-n_j)!}{(k_a-n_i)!(k_b-n_j)!(k_a-n_j)!(k_b-n_1)!} \\
> \ & (1-b_s)\left(\frac{(k_a-(n_j+1))!(k_b-(n_i-1))! + (k_a-(n_i-1))!(k_b-(n_j+1))!}{(k_a-(n_i-1))!(k_b-(n_j+1))!(k_a-(n_j+1))!(k_b-(n_i-1))!}\right)
\end{aligned}$$

$$\begin{aligned}
\Longleftrightarrow \quad & ((k_a-n_j)!(k_b-n_i)! + (k_a-n_i)!(k_b-n_j)!) \\
& ((k_a-(n_i-1))!(k_b-(n_j+1))!(k_a-(n_j+1))!(k_b-(n_i-1))!) \\
> \ & (1-b_s)\left(((k_a-(n_j+1))!(k_b-(n_i-1))! + (k_a-(n_i-1))!(k_b-(n_j+1))!)\right. \\
& \left.((k_a-n_i)!(k_b-n_j)!(k_a-n_j)!(k_b-n_i)!)\right)
\end{aligned}$$

$$\begin{aligned}
\Longleftrightarrow \quad & (k_a-(n_i-1))(k_b-(n_i-1))\left((k_a-n_j)!(k_b-n_i)! + (k_a-n_i)!(k_b-n_j)!\right) \\
> \ & (1-b_s)(k_b-n_j)(k_a-n_j) \\
& ((k_a-(n_j+1))!(k_b-(n_i-1))! + (k_a-(n_i-1))!(k_b-(n_j+1))!) \qquad (2.11)
\end{aligned}$$

We first consider the LHS

$$\begin{aligned}
& (k_a-(n_i-1))(k_b-(n_i-1))\left((k_a-n_j)!(k_b-n_i)! + (k_a-n_i)!(k_b-n_j)!\right) \\
= \ & (k_a-(n_i-1))(k_a-n_j)!(k_b-(n_i-1))! \\
& + (k_a-(n_i-1))!(k_b-(n_i-1))(k_b-n_j)! \\
= \ & (k_a-(n_i-1))!(k_b-(n_i-1))! \\
& \left((k_a-(n_i-1))\left(\prod_{x=n_j}^{n_i-2}(k_a-x)\right) + (k_b-(n_i-1))\left(\prod_{x=n_j}^{n_i-2}(k_b-x)\right)\right),
\end{aligned}$$

and now we consider the RHS

$$\begin{aligned}
& (1-b_s)(k_b-n_j)(k_a-n_j) \\
& ((k_a-(n_j+1))!(k_b-(n_i-1))! + (k_a-(n_i-1))!(k_b-(n_j+1))!) \\
= \; & (1-b_s)\left((k_a-n_j)!(k_b-n_j)(k_b-(n_i-1))!\right. \\
& \left. + (k_a-n_j)(k_a-(n_i-1))!(k_b-n_j)!\right) \\
= \; & (1-b_s)(k_a-(n_i-1))!(k_b-(n_i-1))! \\
& \left((k_b-n_j)\left(\prod_{x=n_j}^{n_i-2}(k_a-x)\right) + (k_a-n_j)\left(\prod_{x=n_j}^{n_i-2}(k_b-x)\right)\right).
\end{aligned}$$

Taking out the common factor of $(k_a-(n_i-1))!(k_b-(n_i-1))!$ means that we can re-write (2.11) as

$$\begin{aligned}
& (k_a-(n_i-1))\left(\prod_{x=n_j}^{n_i-2}(k_a-x)\right) + (k_b-(n_i-1))\left(\prod_{x=n_j}^{n_i-2}(k_b-x)\right) \\
> \; & (1-b_s)\left((k_b-n_j)\left(\prod_{x=n_j}^{n_i-2}(k_a-x)\right) + (k_a-n_j)\left(\prod_{x=n_j}^{n_i-2}(k_b-x)\right)\right) \\
\iff \; & ((k_a-(n_i-1))-(1-b_s)(k_b-n_j))\prod_{x=n_j}^{n_i-2}(k_a-x) \\
> \; & ((1-b_s)(k_a-n_j)-(k_b-(n_i-1)))\prod_{x=n_j}^{n_i-2}(k_b-x), \qquad (2.12)
\end{aligned}$$

i.e. provided that we have

$$(k_a-(n_i-1))-(1-b_s)(k_b-n_j) > 0, \qquad (2.13)$$

we have that (2.12) is equivalent to

$$\frac{\prod_{x=n_j}^{n_i-2}(k_a-x)}{\prod_{x=n_j}^{n_i-2}(k_b-x)} > \frac{(1-b_s)(k_a-n_j)-(k_b-(n_i-1))}{(k_a-(n_i-1))-(1-b_s)(k_b-n_j)}. \qquad (2.14)$$

If we set $b_s = \frac{\max\{n_i\}}{s} \geq \frac{n_i}{k_b}$ then the LHS of (2.13) is not less than

$$\begin{aligned}
k_a-n_i+1-\left(1-\frac{n_i}{k_b}\right)(k_b-n_j) &> 0 \\
\iff k_a-n_i+1-k_b+n_j+n_i-\frac{n_i n_j}{k_b} &> 0 \\
\iff (k_a-k_b)+(n_j-\frac{n_i}{k_b}n_j)+1 &> 0,
\end{aligned}$$

and since

$$k_a \geq k_b,\ \frac{n_i}{k_b} < 1,$$

this inequality trivially holds and we have that the derivation given by (2.14) is valid. We can now turn our attention to showing that (2.14) holds. Note that

$$\frac{\prod_{x=n_j}^{n_i-2}(k_a - x)}{\prod_{x=n_j}^{n_i-2}(k_b - x)} \geq 1,$$

so if we can show that

$$\frac{(1-b_s)(k_a - n_j) - (k_b - (n_i - 1))}{(k_a - (n_i - 1)) - (1-b_s)(k_b - n_j)} < 1, \tag{2.15}$$

then we will have that (2.14) holds. We again take $b_s = \frac{\max\{n_i\}}{s} \geq \frac{n_i}{k_b}$, giving that the LHS of (2.15) $\leq$

$$\begin{aligned}
\frac{\left(1-\frac{n_i}{k_b}\right)(k_a - n_j) - (k_b - (n_i-1))}{(k_a - (n_i-1)) - \left(1-\frac{n_i}{k_b}\right)(k_b - n_j)} &< 1 \\
\iff \left(1-\frac{n_i}{k_b}\right)(k_a - n_j) - (k_b - (n_i-1)) &< (k_a - (n_i-1)) - \left(1-\frac{n_i}{k_b}\right)(k_b - n_j) \\
\iff k_a - n_j - \frac{n_i k_a}{k_b} - \frac{n_i n_j}{k_b} - k_b + n_i - 1 &< k_a - n_i + 1 - k_b + n_j + n_i - \frac{n_i n_j}{k_b} \\
\iff -n_j - \frac{n_i k_a}{k_b} + n_i - 1 &< 1 + n_j \\
\iff n_i - \frac{k_a}{k_b} n_i &< 2 + 2n_j,
\end{aligned}$$

and since we have that

$$n_i - \frac{k_a}{k_b} n_i \leq 0,\ 2 + 2n_j > 0,$$

we have that this inequality holds, as required. Since we picked our permutation τ_{ab} arbitrarily we have that this value of b_s will, for every pair of values a, b, give that

$$\begin{aligned}
& F'\left(\langle n_{\tau_{ab}(1)}, \ldots, n_i, \ldots, n_j, \ldots, n_{\tau_{ab}(2^p)}\rangle, \langle k_1, \ldots, k_{2^p}\rangle\right) \\
& \qquad + F'\left(\langle n_{\overline{\tau}_{ab}(1)}, \ldots, n_j, \ldots, n_i, \ldots, n_{\overline{\tau}_{ab}(2^p)}\rangle, \langle k_1, \ldots, k_{2^p}\rangle\right) \\
> \ & (1-b_s)\left(F'\left(\langle n_{\tau_{ab}(1)}, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_{\tau_{ab}(2^p)}\rangle, \langle k_1, \ldots, k_{2^p}\rangle\right)\right. \\
& \qquad \left. + F'\left(\langle n_{\overline{\tau}_{ab}(1)}, \ldots, n_j + 1, \ldots, n_i - 1, \ldots, n_{\overline{\tau}_{ab}(2^p)}\rangle, \langle k_1, \ldots, k_{2^p}\rangle\right)\right),
\end{aligned} \tag{2.16}$$

and therefore we have that (2.9) holds and, since (2.9) is equivalent to (2.8), we have found the necessary b_s, as required.

We now look at the case where some k_i's are equal to 0. We may assume that

$$N_0 = |\{l \mid n_l = 0\}| \ \geq \ |\{l \mid k_l = 0\}| = K_0$$

since otherwise $F(\vec{\theta}, \vec{\psi}) = F(\vec{\phi}, \vec{\psi}) = 0$ and the result holds trivially.

If $N_0 = K_0$ then we may take that $n_j \neq 0$ since otherwise

$$F(\vec{\phi}, \vec{\psi}) = 0,$$

and the result holds trivially. Taking $N_0 = K_0$, $n_j > 0$ gives that

$$\begin{aligned} & F(\vec{\theta}, \vec{\psi}) > (1 - b_s)F(\vec{\phi}, \vec{\psi}) \\ \iff \quad & F(\vec{\theta'}, \vec{\psi'}) > (1 - b_s)F(\vec{\phi'}, \vec{\psi'}) \end{aligned}$$

where θ', ϕ', ψ' are the non-zero $2^p - K_0$ terms of θ, ϕ, ψ respectively. The only permutations which don't give rise to $F'(\vec{\theta}, \hat{\psi}) = F'(\vec{\phi}, \hat{\psi}) = 0$ are the ones where the zeros are in positions occupied by zeros in $\vec{k}$. The result then follows by the same proof as before.

When $N_0 > K_0$ we have that

$$\begin{aligned} & F(\vec{\theta}, \vec{\psi}) > (1 - b_s)F(\vec{\phi}, \vec{\psi}) \\ \iff \quad & C_\theta \cdot F(\vec{\theta'}, \vec{\psi'}) > C_\phi \cdot (1 - b_s)F(\vec{\phi'}, \vec{\psi'}), \end{aligned} \tag{2.17}$$

where, when $n_j > 0$,

$$C_\theta = C_\phi = \frac{N_0!}{(N_0 - K_0)!K_0!}$$

since it matters which zero coordinates we choose to include in θ', ϕ' but not the order in which we include them. When $n_j = 0$

$$C_\theta = \frac{N_0!}{(N_0 - K_0)!K_0!}, \quad C_\phi = \frac{(N_0 - 1)!}{(N_0 - 1 - K_0)!K_0!}$$

because $m_j = 1$ must be included in $\vec{\phi}'$ meaning that there is one less position to "give" to one of the zero coordinates. Since this last case gives

$$\frac{C_\theta}{C_\phi} = \frac{N_0}{N_0 - K_0},$$

and $K_0 > 0$, we have that

$$C_\theta > C_\phi,$$

in all cases, and so (2.17) holds, and therefore so to does the result, as required.

All cases are now covered and, as such, the result is shown. $\square$

Lemma 2.4.4. *Take $s \gg \max n_i$, then*

$$\sum_{\hat{\psi} \in D_s^K} F(\vec{\theta}, \vec{\psi}) w(\psi) \to_{K \to \infty} 0,$$

where $D_s^K = \left\{ \vec{k} \mid (\sum_i k_i = K) \wedge (\{i \mid 0 < k_i < s\} \neq \emptyset) \wedge (k_1 \geq k_2 \geq \ldots \geq k_{2^p}) \right\}$.

Proof. To prove this it will be enough to show that

$$\sum_{\hat{\psi} \in D_s^K} F(\emptyset, \vec{\psi}) w(\psi) \to_{K \to \infty} 0, \tag{2.18}$$

since we trivially have that

$$\sum_{\hat{\psi} \in D_s^K} F(\vec{\theta}, \vec{\psi}) w(\psi) \leq \sum_{\hat{\psi} \in D_s^K} F(\emptyset, \vec{\psi}) w(\psi).$$

Let

$$H = \{i \mid 0 \leq k_i < s\},$$

and

$$G_r^K = \left\{ \vec{k} \mid \left(\sum_i k_i = K \right) \wedge \left(\sum_{i \in H} k_i \leq r \right) \wedge (k_1 \geq k_2 \geq \ldots \geq k_{2^p}) \right\},$$

then a limit, as $K \to \infty$, does exist for

$$\sum_{\hat{\psi} \in G_r^K} F(\emptyset, \vec{\psi}) w(\psi),$$

since we have that

$$\sum_{\hat{\psi} \in G_r^K} F(\emptyset, \vec{\psi}) w(\psi) \geq \sum_{\hat{\psi} \in G_r^{K+1}} F(\emptyset, \vec{\psi}) w(\psi),$$

as all the ψ such that $\hat{\psi} \in G_r^{K+1}$ are extensions of the ψ where $\hat{\psi} \in G_r^K$ but not all such extensions are in G_r^{K+1}. Let this limit be denoted by γ_r. If $\gamma_r = \gamma_{r-1}$ for all $r > 0$ then (2.18) will hold because

$$D_s^K \subset \bigcup_{r>0} \left(G_r^K - G_{r-1}^K \right),$$

indeed,

$$D_s^K \subset \bigcup_{r=1}^{2^p s} \left(G_r^K - G_{r-1}^K \right).$$

Suppose that we do not have $\gamma_r = \gamma_{r-1}$ and that r is minimal such that this fails. Then we must have that $\gamma_r > \gamma_{r-1}$, say $\gamma_r - \gamma_{r-1} = \gamma > 0$. Hence

$$\lim_{K \to \infty} \sum_{\hat{\psi} \in G_r^K - G_{r-1}^K} F(\emptyset, \vec{\psi}) w(\psi) = \gamma_r - \gamma_{r-1} = \gamma > 0. \tag{2.19}$$

Let A_r^K denote the set of vectors $\langle q_1, q_2, \ldots, q_K \rangle$ where

$$\sum_{\substack{1 \leq v \leq 2^p \\ |\{j | q_j = v\}| \leq s}} |\{j \mid q_j = v\}| = r,$$

i.e. these represent the possible streams of observations for the state descriptions whose ordered vector representations will appear in $G_r^K - G_{r-1}^K$. We therefore have that

$$\lim_{K \to \infty} \sum_{\vec{q} \in A_r^K} w \left(\bigwedge_{j=1}^K \alpha_{q_j}(a_j) \right) = \gamma,$$

is equivalent to (2.19). Let

$$A_r'^K = \bigcup_{r' < r} A_{r'}^K,$$

i.e. the list of possible streams of observations that are such that their ordered vector representation will be in the set G_{r-1}^K, let K be large enough such that

$$\sum_{\vec{q} \in A_r'^K} w \left(\bigwedge_{j=1}^K \alpha_{q_j}(a_j) \right) < \gamma_{r-1} + \frac{\gamma}{3}.$$

Now pick T large enough such that

$$\sum_{\vec{q}\in A_r^{KT}} w\left(\bigwedge_{j=1}^{KT}\alpha_{q_j}(a_j)\right) > \gamma_r - \frac{\gamma}{3}.$$

Let E_r^{KT} be the set of those $\langle q_1,\ldots,q_{KT}\rangle$ for which $\langle q_1,\ldots,q_K\rangle \in A_r^K$, i.e. where all occurrences of the scarce atoms have occurred before the $K+1$'st observation. Since $A_r^{KT} - E_r^{KT}$ only contains extensions of the state descriptions in $A_r'^K$, but not all of them, we must have that

$$\begin{aligned}\sum_{\vec{q}\in A_r^{KT}-E_r^{KT}} w\left(\bigwedge_{j=1}^{KT}\alpha_{q_j}(a_j)\right) &\leq \sum_{\vec{q}\in A_r'^K} w\left(\bigwedge_{j=1}^{K}\alpha_{q_j}(a_j)\right)\\ &< \gamma_{r-1} + \frac{\gamma}{3},\end{aligned}$$

so

$$\sum_{\vec{q}\in E_r^{KT}} w\left(\bigwedge_{j=1}^{KT}\alpha_{q_j}(a_j)\right) > \gamma_r - \frac{\gamma}{3} - \left(\gamma_{r-1}+\frac{\gamma_r}{3}\right) = \gamma - \frac{\gamma}{3} - \frac{\gamma}{3} = \frac{\gamma}{3}. \tag{2.20}$$

For $0 \leq v < T$ let ψ_v be the result of simultaneously transposing a_j with a_{Kv+j} in

$$\bigvee_{\vec{q}\in E_r^{KT}} \bigwedge_{j=1}^{KT}\alpha_{q_j}(a_j),$$

for $j = 1, 2, \ldots, K$. By (2.20) and Ex each $w(\psi_v) > \frac{\gamma}{3}$ for $0 \leq v \leq T$. But these ψ_v are disjoint so

$$w\left(\bigvee_{v=0}^{T-1}\psi_v\right) = \sum_{v=0}^{T-1} w(\psi_v) > \frac{T\gamma}{3}.$$

But we can choose T so large that $\frac{T\gamma}{3} > 1$, which thereby forms the required contradiction. □

We can now take these lemmas and use them to give our "Long Road" proof to Theorem 2.4.1 which, since it has been several pages since we last saw it, we will now restate.

Theorem 2.4.1 *If we have θ, ϕ as defined above (i.e. where they differ by one observation and are not exchangeable) then for any probability function satisfying Ax we have that*

$$w(\theta) \geq w(\phi). \tag{2.21}$$

Proof. For any given large K we have that

$$w(\theta) = \sum_{\hat{\psi} \in Q_K} F\left(\vec{\theta}, \vec{\psi}\right) \cdot w(\psi),$$

where

$$Q_K = \left\{\vec{k} \mid \left(\sum_i k_i = K\right) \wedge (k_1 \geq k_2 \geq \ldots k_{2^p})\right\}$$

. With D_s^k, B_s^K as defined in our previous two lemmas we have that

$$w(\theta) = \sum_{\hat{\psi} \in B_s^K} F\left(\vec{\theta}, \vec{\psi}\right) \cdot w(\psi) + \sum_{\hat{\psi} \in D_s^K} F\left(\vec{\theta}, \vec{\psi}\right) \cdot w(\psi).$$

Pick $s \gg \max n_i$ so that, for large K,

$$\begin{aligned} w(\theta) &\geq \sum_{\hat{\psi} \in B_s^K} F\left(\vec{\theta}, \vec{\psi}\right) \cdot w(\psi) \\ &\geq (1 - b_s) \sum_{\hat{\psi} \in B_s^K} F\left(\vec{\phi}, \vec{\psi}\right) \cdot w(\psi). \end{aligned} \tag{2.22}$$

Because

$$w(\phi) = \sum_{\hat{\psi} \in B_s^K} F\left(\vec{\phi}, \vec{\psi}\right) \cdot w(\psi) + \sum_{\hat{\psi} \in D_s^K} F\left(\vec{\phi}, \vec{\psi}\right) \cdot w(\psi),$$

and

$$\sum_{\hat{\psi} \in D_s^K} F\left(\vec{\phi}, \vec{\psi}\right) \cdot w(\psi) \to_{K \to \infty} 0,$$

by Lemma 2.4.4. Given $\epsilon > 0$, we have from (2.22) that, for large K

$$\begin{aligned} w(\theta) &\geq (1 - b_s) w(\phi) - \sum_{\hat{\psi} \in D_s^K} F\left(\vec{\phi}, \vec{\psi}\right) \cdot w(\psi) \\ &\geq (1 - b_s) w(\phi) - \epsilon. \end{aligned}$$

Since we can choose ϵ and b_s to be as small as we'd like, we have that

$$w(\theta) \geq w(\phi),$$

as required. □

2.4.3 The Only Rule

We now move on to proving the main theorem of the chapter. We start with a couple of definitions and a lemma which will provide the framework for the proof. Let θ, ϕ be as defined in Section 2.2, with *ordered* vector representations given by

$$\hat{\theta} = \langle n_1, n_2, \dots, n_{2^p} \rangle,$$

$$\hat{\phi} = \langle m_1, m_2, \dots, m_{2^p} \rangle,$$

Definition 2.4.5. Given a vector $\hat{\theta}$ as defined above and $1 \leq i \leq 2^p - 1,\ 2 \leq j \leq 2^p$, $n_i > n_j + 1$, i is maximal and j is minimal amongst elements of the same magnitude, an *m-step* is the process which transforms $\hat{\theta}$ into the vector $\hat{\theta}' = \langle n_1, \dots, n_i - 1, \dots, n_j + 1, \dots, n_{2^p} \rangle$, and is denoted $\hat{\theta} \rightarrow_m \hat{\theta}'$.

Note that by Theorem 2.4.1 we have that $w(\hat{\theta}) \geq w(\hat{\theta}')$. This concept allows us a simple way to describe an application of Theorem 2.4.1, and also allows us to bring in the concept of a "route":

Definition 2.4.6. Given θ and ϕ, as defined earlier, an *m-route*, is defined to exist from $\hat{\theta}$ to $\hat{\phi}$ if either the two vectors are equal (a null m-route) or if it is possible to get from $\hat{\theta}$ to $\hat{\phi}$ via a series of m-steps.
Let $\hat{\theta} \Rightarrow_m \hat{\phi}$ denote that there exists an m-route between $\hat{\theta}$ and $\hat{\phi}$ and $\hat{\theta} \not\Rightarrow_m \hat{\phi}$ denote that such an m-route does not exist.
The $\hat{\theta}$ here is known as the initial vector whilst $\hat{\phi}$ is called the destination vector.

We can immediately state and prove a simple lemma regarding these routes.

Lemma 2.4.7. *Let θ, ϕ, be as defined above then we have that*

$$\left(\hat{\theta} \Rightarrow_m \hat{\phi}\right) \Rightarrow w(\theta) \geq w(\phi)$$

Proof. Follows immediately from Theorem 2.4.1 and the definitions of an m-step and an m-route. □

We can now state and prove The Only Rule.

Theorem 2.4.8. *For $\theta, \phi \in SL$ as defined above, a necessary and sufficient condition for*

$$w(\theta) \geq w(\phi), \tag{2.23}$$

to hold for all w satisfying Ax is that

$$\sum_{j \geq i} m_j \geq \sum_{j \geq i} n_j \quad \textit{for } i = 1, 2, 3 \ldots 2^p. \tag{2.24}$$

Proof. We first show that if we don't have (2.24) then we can find some probability function satisfying Ax such that we also do not have (2.23).

If we do not have (2.24) for $i = 1$ (i.e. there are more observations in θ than in ϕ, so $N > M$) then the Independent Solution[2], w_I gives

$$w_I(\theta) = \left(\frac{1}{2^p}\right)^N < \left(\frac{1}{2^p}\right)^M = w_I(\phi),$$

and as such provides a counter-example to (2.23).

If (2.24) fails for some $i > 1$ then let w be the SM function based upon the measure which puts weight 1 on the point $\vec{b} \in \mathbb{D}_{2^p}$, where $\vec{b}$ is the vector that has the first $i-1$ coordinates equal to $\frac{1-(i-1)\epsilon}{2^k-(i-1)}$ and the rest equal to ϵ for very small ϵ. Then, by the definition of SM functions, we have that the probabilities of θ and ϕ will be given by

$$2^p! \cdot w(\phi) = \sum_{\tau} \left(\frac{1-(i-1)\epsilon}{2^k-(i-1)}\right)^{\sum_{j<i} m_{\tau(j)}} \epsilon^{\sum_{j \geq i} m_{\tau(j)}}, \tag{2.25}$$

$$2^p! \cdot w(\theta) = \sum_{\tau} \left(\frac{1-(i-1)\epsilon}{2^k-(i-1)}\right)^{\sum_{j<i} n_{\tau(j)}} \epsilon^{\sum_{j \geq i} n_{\tau(j)}}, \tag{2.26}$$

where τ ranges over all permutations of $\{1, 2, \ldots, 2^p\}$. Due to the ordering of the terms in $\hat{\theta_\sigma}$ we have that the smallest power of ϵ in (2.26) is given by $\sum_{j \geq i} n_j$. The corresponding power of ϵ in (2.25) is given by $\sum_{j \geq i} m_j$ which is less than $\sum_{j \geq i} n_j$ because (2.24) fails at this point. Because we can take ϵ to be as small as we like we

[2] See page 23 for the definition of the Independent Solution.

must have that

$$\sum_{\tau} \left(\frac{1-(i-1)\epsilon}{2^k-(i-1)} \right)^{\sum_{j<i} m_{\tau(j)}} \epsilon^{\sum_{j \geq i} m_{\tau(j)}}$$
$$> \sum_{\tau} \left(\frac{1-(i-1)\epsilon}{2^k-(i-1)} \right)^{\sum_{j<i} n_{\tau(j)}} \epsilon^{\sum_{j \geq i} n_{\tau(j)}},$$

i.e.

$$w(\phi) > w(\theta),$$

giving a counter-example to (2.23), as required.

Now suppose that (2.24) holds. We may assume that $N = M$ since otherwise $M > N$ and we can repeatedly remove 1 from m_r where (at each step) r is maximal such that

$$\sum_{j \geq i} m_j > \sum_{j \geq i} n_j, \tag{2.27}$$

for all $i \leq r$, until $\hat{\phi}'_\sigma$ contains N observations. It is always possible to find such a r since $r = 1$ will satisfy (2.27) whenever $M > N$. This new vector $\vec{\phi}'_\sigma$ will still satisfy (2.24) and, because ϕ represents some extension of ϕ', has the property that

$$w(\phi) \leq w(\phi').$$

for any w.

We now aim to show that when we have $N = M$ and (2.24) we also have that $\hat{\theta} \Rightarrow_m \hat{\phi}$ and the result will then follow from Lemma 2.4.7.

Let $\vec{r}$ to be the vector with components given by

$$r_i = \sum_{j \geq i} m_j - \sum_{j \geq i} n_j.$$

We call this the running difference vector between $\hat{\phi}$ and $\hat{\theta}$. Note that the condition given by (2.24) is equivalent to the running difference vector containing no negative components. Let $R = \sum_i r_i$.

If $R = 0$ then we have that the two vectors must be equal, since each $r_i = 0$, and this gives that we have a (null) m-route between the two vectors as required. We now consider the case when $R > 0$. We provide an algorithm which gives an m-route between $\hat{\theta}$ and $\hat{\phi}$ as required.

Step 1 : Take maximal s such that r_s is the maximum value over all the r_i and maximal $t < s$ such that $r_t < r_s$.
When $R > 0$ it will trivially always be possible to find such an s, it will always be possible to find such a t because when $N = M$ we must have that $r_1 = 0 < r_s$. This choice means that $n_t > m_t \geq m_s > n_s$, indeed it means that for each $t < j < s$,

$$n_t > m_t \geq m_j = n_j \geq m_s > n_s.$$

Step 2 : Perform an m-step on $\hat{\theta}$, decreasing n_t and increasing n_s by one. Note that all our conditions are met, particularly that $n_t > n_s + 1$ and that t, s are minimal/maximal, respectively, amongst components of the same magnitude (these ensure that our destination vector is also an ordered vector representation).
Let the destination vector be called $\hat{\theta}'$ and the running difference vector between $\hat{\phi}$ and $\hat{\theta}'$ be called $\vec{r'}$.

When $i \leq t$ we have that $r'_i = r_i$ because:

$$\begin{aligned}
r'_i &= \sum_{j\geq i} m_j - \sum_{j\geq i} n'_j \\
&= \sum_{j\geq i} m_j - (n_i + n_{i+1} + \ldots + (n_t - 1) + \ldots + (n_s + 1) + \ldots + n_{2^p}) \\
&= \sum_{j\geq i} m_j - (n_i + n_{i+1} + \ldots + n_t + \ldots + n_s + \ldots + n_{2^p} + 1 - 1) \\
&= \sum_{j\geq i} m_j - \sum_{j\geq i} n_j \\
&= r_i.
\end{aligned}$$

When $t < i \leq s$ we have that $r'_i = r_i - 1$ because

$$\begin{aligned} r'_i &= \sum_{j \geq i} m_j - \sum_{j \geq i} n'_j \\ &= \sum_{j \geq i} m_j - (n_i + n_{i+1} + \ldots + (n_s + 1) + \ldots + n_{2^p}) \\ &= \sum_{j \geq i} m_j - (n_i + n_{i+1} + \ldots + n_t + \ldots + n_s + \ldots + n_{2^p}) - 1 \\ &= \sum_{j \geq i} m_j - \sum_{j \geq i} n_j - 1 \\ &= r_i - 1. \end{aligned}$$

and when $s < i$ we again have that $r'_i = r_i$, because

$$\begin{aligned} r'_i &= \sum_{j \geq i} m_j - \sum_{j \geq i} n'_j \\ &= \sum_{j \geq i} m_j - (n_i + n_{i+1} + \ldots + n_{2^p}) \\ &= \sum_{j \geq i} m_j - \sum_{j \geq i} n_j \\ &= r_i. \end{aligned}$$

Due to the way in which we selected n_t and n_s the r_i which decrease by 1 must be greater than 0, therefore our new $\vec{r'}$ will have the property that every $r'_i \geq 0$.

Step 3 : If $R' = \sum_i r'_i = 0$ then we stop because $\hat{\theta}' = \hat{\phi}$ (since every r'_i must equal 0) and we will have successfully found our m-route between $\hat{\theta}$ and $\hat{\phi}$. Otherwise go back to step 1, setting $\theta = \theta'$ $\vec{r} = \vec{r'}$, $(R = R')$ to find our next m-step.

The algorithm must complete because if $R > 0$ we can always find a next step to perform which will reduce R (since $t < s$ and $r'_i = r_i - 1$ in the region $t < i \leq s$) without making $R < 0$ (since the new running difference vector we form cannot contain any negative coordinates) and therefore we can always find an m-route between $\hat{\theta}$ and $\hat{\phi}$, as required. Therefore this direction is complete.

$\square$

Note that The Only Rule states that the more "balanced" a state description is, the lower its probability will be or, looking at it the other way around, the highest probabilities are awarded to the homogenous state descriptions. So we have that

$$\begin{aligned} & w(\langle n_1, 0, \ldots, 0\rangle) \geq w(\langle n_1 - 1, 1, 0, \ldots, 0\rangle) \geq w(\langle n_1 - 2, 2, 0, \ldots, 0\rangle) \\ & \quad \geq w(n_1 - 2, 1, 1, 0, \ldots, 0\rangle) \geq \ldots \geq w(\langle \left[\frac{n_1}{2^p}\right] + c_1, \left[\frac{n_1}{2^p}\right] + c_2, \ldots, \left[\frac{n_1}{2^p}\right]\rangle), \end{aligned}$$

where

$$c_i = \begin{cases} 1 & i \leq \text{ the remainder when } n_1 \text{ is divided by } 2^p, \\ 0 & \text{otherwise.} \end{cases}$$

Chapter 3

Principles Involving Relevance

3.1 Introduction & Motivation

In this chapter we concern ourselves with principles based upon the idea of relevance, that is to say what we should be expecting based upon what we have witnessed in the past. The obvious starting point is the famous Principle of Instantial Relevance[1] (PIR), introduced by Carnap in [3], and defined as follows.

Definition 3.1.1. For an atom $\alpha(x)$ and state description $\phi(a_1, a_2, \ldots, a_n)$ the *Principle of Instantial Relevance* is the requirement that

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)).$$

In essence, then, this principle states that

> "If we have observed $\phi(a_1, a_2, \ldots a_n)$ then our belief in $\alpha(a_{n+2})$ holding should not be decreased if you also learn that $\alpha(a_{n+1})$ holds, i.e. learning that a_{n+1} satisfies α should (if anything) enhance your belief that a_{n+2} will too."

PIR can be shown to follow directly from Ex via de Finetti's theorem (see, for example [19]), thus confirming the apparent reasonableness (and therefore desirability) of the principle.

[1]This is defined earlier, on page 18, but due to the prominent role the principle plays throughout this chapter we have decided to also include it here.

In [23], Chris Nix introduced a generalised version of PIR, GPIR (defined here on page 19), which turns out to be much stronger than PIR, satisfying only the narrow range of probability functions known as the Paris-Nix Continuum. Here we introduce two new principles which are also related to PIR, the Strong Principle of Instantial Relevance (SPIR) and the Negative Principle of Instantial Relevance (NPIR).

The first part of our motivation for this chapter can be summed up as

"What can we say about where our two new principles, SPIR & NPIR, hold with reference to symmetric probability functions?"

NPIR differs from PIR by stating what we should expect if, instead of observing an extra instance of $\alpha(x)$, we'd observed an extra instance of $\beta(x)$, where $\beta(x)$ is some other atom. SPIR takes this concept one further by stating how we should expect the probabilities of observing $\alpha(x)$ next should compare if, on one hand we had observed an extra $\alpha(x)$, and on the other an extra $\beta(x)$. Both are explained in more detail in the next section.

We start our investigation into these two principles by giving a rather obvious result stating how the two principles are related to each other and PIR. We then proceed by showing that both are satisfied by the continuum's of Carnap and Paris-Nix to the extent where both are satisfied strictly in all non-extreme cases.

We then widen our focus to include all probability functions that satisfy Ax. Here we start by giving some results showing where NPIR & SPIR hold for all such probability functions, in terms of the number of predicates in the given language and then related to the number of observations made.

We will then give some rather surprising results regarding situations in which SPIR and NPIR do not hold. This section will include some generalised counter-examples and analysis of the results to try and find out exactly where we can say SPIR and NPIR do not hold.

We sum up the overall situation with SPIR, including pointing out where the situation is unresolved, though we then clear up all the unresolved cases for NPIR. Briefly, we consider SPIR's (and by implication, NPIR's) relationship to Ex.

We finish this part of the chapter with a brief discussion based upon what we have found.

In the final section of the chapter we return to look specifically at PIR.

As we have already mentioned, de Finetti's theorem provides a neat proof for $Ex \Rightarrow PIR$. Our motivation here is to consider the previously unstudied converse, i.e.

"Does PIR imply Ex?"

We discover that this implication fails for what we would regard to be a sensible version of PIR, indeed a version which includes the version we use throughout the rest of the thesis. However, when we consider our counter-example with relation to the most general form of PIR (not to be confused with GPIR) we find that it no longer holds, and that the situation here is yet to be resolved.

3.2 Some Preliminaries

We now introduce our new principles, starting with the Negative Principle of Instantial Relevance defined as follows;

Definition 3.2.1. For (distinct) atoms $\alpha(x), \beta(x)$ and state description $\phi(a_1, a_2, \ldots, a_n)$ the *Negative Principle of Instantial Relevance* (NPIR) is the requirement that

$$w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)).$$

That is to say

"If we have observed $\phi(a_1, a_2, \dots a_n)$ then our belief in $\alpha(a_{n+2})$ holding should not be increased if you also learn that a_{n+1} satisfies some other atom, β."

The Strong Principle of Instantial Relevance is defined as follows;

Definition 3.2.2. For (distinct) atoms $\alpha(x), \beta(x)$ and state description $\phi(a_1, a_2, \dots, a_n)$ the *Strong Principle of Instantial Relevance* (SPIR) is the requirement that

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \dots, a_n)) \geq w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \dots, a_n)).$$

This essentially says

"In addition to knowing $\phi(a_1, a_2, \dots a_n)$, if we learn that a_{n+1} satisfies α our belief that a_{n+2} does too should be higher than if we learn that a_{n+1} satisfies some other atom, β."

In these definitions we take

$$w(\theta_1 \mid \psi_1) \geq w(\theta_2 \mid \psi_2),$$

to mean

$$w(\theta_1 \wedge \psi_1) \cdot w(\psi_2) \geq w(\theta_2 \wedge \psi_2) \cdot w(\psi_1).$$

These principles ostensibly appear to be desirable and sensible, indeed they are clearly both closely related to PIR, but things are not always as straightforward as they at first appear!

Initially we will study just the areas where these principles do act as we would expect them to, but before we continue we provide a theorem which links together PIR, SPIR and NPIR, mathematically.

Theorem 3.2.3. *In the presence of PIR, NPIR implies SPIR, that is to say;*

$$PIR \ \&\ NPIR \Rightarrow SPIR.$$

Further, if we have the strict version of NPIR, where the $\geq$ is replaced by $>$, we have that in the presence of PIR, NPIR implies the strict version of SPIR.

Proof. Let $\alpha(x), \beta(x)$ be states and $\phi(a_1, a_2 \ldots, a_n)$ some state description for a given language L. Then, by PIR, we have

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)), \tag{3.1}$$

and by NPIR we have

$$w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)). \tag{3.2}$$

Putting (3.1) and (3.2) together we have that

$$\begin{aligned} w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) &\geq w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) \\ &\geq w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)), \end{aligned}$$

i.e.

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)), \tag{3.3}$$

which is SPIR, as required.

Note that if we replace the $\geq$ with $>$ in (3.2) then (3.3) becomes

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) > w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)).$$

So the strict version of NPIR implies the strict version of SPIR, as required. □

A trivial corollary that we will be using later is now given.

Corollary 3.2.4. *In the presence of PIR, if we do not have SPIR then we do not have NPIR, i.e.*

$$PIR \& \neg SPIR \Rightarrow \neg NPIR$$

Proof. Follows logically from Theorem 3.2.3. □

A final useful theorem before we continue.

Theorem 3.2.5. *Where $\alpha(x), \beta(x), \phi(\vec{x})$ are defined as in the definitions for NPIR & SPIR, we have that both principles hold with equality when we have that*

(a) $w(\phi(a_1, \dots a_n)) = 0,$

(b) $w(\alpha(a_{n+2}) \wedge \phi(a_1, \dots, a_n)) = 0$ *or*

(c) $w(\beta(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) = 0.$

Proof. First note that for any sentences $\psi_1, \psi_2 \in SL$, if we have that

$$w(\psi_1) = 0,$$

then we also have that

$$w(\psi_1 \wedge \psi_2) = 0.$$

We will be using this fact often, and without further mention, throughout this proof.

We consider NPIR to be defined as

$$\begin{aligned} & w(\alpha(a_{n+2}) \wedge \phi(a_1, \dots, a_n)) \cdot w(\beta(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) \\ \geq \; & w(\alpha(a_{n+2}) \wedge \beta(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) \cdot w(\phi(a_1, \dots, a_n)), \end{aligned}$$

and SPIR as

$$\begin{aligned} & w(\alpha(a_{n+2}) \wedge \alpha(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) \cdot w(\beta(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) \\ \geq \; & w(\alpha(a_{n+2}) \wedge \beta(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) \cdot w(\alpha(a_{n+1}) \wedge \phi(a_1, \dots, a_n)). \end{aligned}$$

For case (a), let

$$w(\phi(a_1, \dots, a_n)) = 0,$$

then trivially the RHS is 0, but also

$$w(\alpha(a_{n+2}) \wedge \phi(a_1, \dots, a_n)) = 0, \; w(\beta(a_{n+1}) \wedge \phi(a_1, \dots, a_n)) = 0, \tag{3.4}$$

so the LHS is 0 too, and therefore NPIR holds with equality.

For SPIR the result follows trivially from (3.4).

For case (b), let

$$w(\alpha(a_{n+2}) \wedge \phi(a_1, \ldots, a_n)) = 0,$$

then trivially the LHS is 0, but also

$$w(\alpha(a_{n+2}) \wedge \beta(a_{n+1}) \wedge \phi(a_1, \ldots, a_n)) = 0, \tag{3.5}$$

so the RHS is 0 too, and therefore NPIR holds with equality.
For SPIR we need the additional fact that

$$w(\alpha(a_{n+2}) \wedge \alpha(a_{n+1}) \wedge \phi(a_1, \ldots, a_n)) = 0$$

and the result then follows trivially from this and (3.5).

Case (c) is similar to case (b), and so all three cases are proven, as required. □

In the sequel, then, we will assume that $w(\phi(a_1, a_2, \ldots, a_n)), w(\alpha(a_{n+2}) \wedge \phi(a_1, a_2, \ldots, a_n))$ and $w(\beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) \neq 0$.

3.3 NPIR and SPIR Hold for the Carnap and Paris-Nix Continuums

We start our investigation by showing that both NPIR and SPIR are satisfied by our two main probability function families that satisfy Ax. Indeed, for all non-extreme values we'll show that both continuums satisfy the strict versions of each principle.

Theorem 3.3.1. *Carnap's Continuum satisfies NPIR & SPIR. Further, unless we take the probability functions formed at the limits $\lambda \to \infty$ (giving the independent solution, w_I) and $\lambda = 0$ (giving Carnap's $\mathfrak{m}_\circ$), the continuum satisfies both principles strictly.*

Proof. By theorem 3.2.3, we only need to show that the continuum satisfies NPIR. Let L be a language containing p predicates, let $\alpha(x), \beta(x)$ be states and $\phi(a_1, a_2, \ldots, a_n)$

a state description for that language. Say ϕ contains $m \geq 0$ observations of $\alpha(x)$ (and n observations in total), then NPIR gives

$$w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)).$$

We first deal with the continuum at its limits. For $\lambda \to \infty$ we have that $w_\lambda = w_I$ which gives[2]

$$w_I(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) = \frac{\left(\frac{1}{2^p}\right)^{n+1}}{\left(\frac{1}{2^p}\right)^{n}} = \frac{1}{2^p},$$

$$w_I(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) = \frac{\left(\frac{1}{2^p}\right)^{n+2}}{\left(\frac{1}{2^p}\right)^{n+1}} = \frac{1}{2^p}$$

and so NPIR and SPIR both hold. In fact both hold with equality, we've already shown that NPIR does so, but by considering

$$w_I(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) = \frac{\left(\frac{1}{2^p}\right)^{n+2}}{\left(\frac{1}{2^p}\right)^{n+1}} = \frac{1}{2^p},$$

it becomes apparent that SPIR does so too.

For $\lambda = 0$ we have the pathological function, $\mathfrak{m}_\circ$. Now NPIR is defined to be

$$w(\alpha \wedge \phi) \cdot w(\beta \wedge \phi) \geq w(\alpha \wedge \beta \wedge \phi) \cdot w(\phi).$$

Since all non-homogenous state descriptions are awarded a probability of 0 under $\mathfrak{m}_\circ$, we must have that either $w_{\mathfrak{m}_\circ}(\alpha \wedge \phi) = 0$ or $w_{\mathfrak{m}_\circ}(\beta \wedge \phi) = 0$ (or both!), so the LHS will always be equal to zero. We also have that $\alpha \wedge \beta \wedge \phi$ defines a non-homogenous state description, so the RHS will also always be equal to zero. In this way NPIR, and similarly SPIR[3], both collapse to triviality here and are satisfied with equality.

Therefore for both limits, NPIR and SPIR hold, as required.

---

[2]The derivation which follows will actually hold for $\lambda \to \infty$, and so it is not strictly necessary to show this case separately. However we do believe that it is useful to show that the Independent Solution satisfies NPIR & SPIR explicitly - partly because it is a standard symmetric probability function in its own right.

[3]Where the only difference is that we require the fact that either $w_{\mathfrak{m}_\circ}(\alpha \wedge \alpha \wedge \phi) = 0$ or $w_{\mathfrak{m}_\circ}(\beta \wedge \phi) = 0$.

We now consider Carnap's Continuum for the range $(0, \infty)$. The definition of w_λ (see page 25) gives that

$$w(\alpha(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) = \frac{m + \frac{\lambda}{2^p}}{n + \lambda}, \tag{3.6}$$

and

$$w(\alpha(a_{n+2}) \mid \beta(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)) = \frac{m + \frac{\lambda}{2^p}}{n + 1 + \lambda}. \tag{3.7}$$

Putting (3.6) and (3.7) together we get that, for NPIR to hold we require

$$\begin{aligned} \frac{m + \frac{\lambda}{2^p}}{n + \lambda} &\geq \frac{m + \frac{\lambda}{2^p}}{n + 1 + \lambda} \\ \iff n + 1 + \lambda &\geq n + \lambda, \end{aligned}$$

and this is trivially true, therefore NPIR & SPIR are satisfied by all probability functions belonging to Carnap's Continuum, as required. Further, we actually have that

$$n + 1 + \lambda > n + \lambda,$$

and working back up gives that the Continuum satisfies NPIR (and therefore SPIR) strictly in the range $(0, \infty)$, as required. □

Theorem 3.3.2. *The Paris-Nix continuum satisfies NPIR & SPIR. Further whenever $\delta \in (0, 1)$ it satisfies them both strictly.*

Proof. By theorem 3.2.3, we only need to show that the continuum satisfies NPIR. Let L be the language with p predicates. Let $\alpha_1(x), \ldots, \alpha_{2^p}(x)$ list the atoms of L. We will henceforth be referring to the Paris-Nix continuum by w^δ. NPIR is defined to be

$$w^\delta(\alpha_1(a_{n+2}) \mid \phi(a_1, a_2, \ldots, a_n)) \geq w^\delta(\alpha_1(a_{n+2}) \mid \alpha_2(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n)),$$

and we'll be dealing specifically with the expanded version given by

$$\frac{w^\delta(\alpha_1(a_{n+2}) \wedge \phi(a_1, a_2, \ldots, a_n))}{w^\delta(\phi(a_1, a_2, \ldots, a_n))} \geq \frac{w^\delta(\alpha_1(a_{n+2}) \wedge \alpha_2(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n))}{w^\delta(\alpha_2(a_{n+1}) \wedge \phi(a_1, a_2, \ldots, a_n))}. \tag{3.8}$$

The following derivation is valid for $\delta \in [0,1)$ as when $\delta = 1$ (3.8) is not always defined (i.e. w^δ does not satisfy REG when $\delta = 1$). However[4], when $\delta = 1$ the Paris-Nix Continuum degenerates to Carnap's $\mathfrak{m}_o$, which we have already shown satisfies NPIR in the proof of Theorem 3.3.1. Let

$$\phi(a_1, a_2, \ldots, a_n) = \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i).$$

We remind ourselves of the definition of w^δ as given on page 26, and apply it to ϕ to give

$$w^\delta(\phi(a_1, a_2, \ldots, a_n)) = \frac{1}{2^p}\sum_{r=1}^{2^p}(\gamma+\delta)^{n_r}\gamma^{n-n_r},$$

where $n_r = |\{i \mid h_i = r\}|$ for $r = 1, 2, \ldots, 2^p$, $2^p\gamma = 1-\delta$ and $\delta \in [0,1)$, in particular note that the number of times α_1 appears in ϕ equals n_1, and similarly for α_2, n_2. Let

$$\chi_0 = \sum_{r=3}^{2^p}(\gamma+\delta)^{n_r}\gamma^{n-n_r},$$

then we have that

$$w^\delta(\alpha_1 \wedge \phi) = \frac{1}{2^p}\left((\gamma+\delta)^{n_1+1}\gamma^{n-n_1} + (\gamma+\delta)^{n_2}\gamma^{n+1-n_2} + \gamma\chi_0\right), \tag{3.9}$$

$$w^\delta(\phi) = \frac{1}{2^p}\left((\gamma+\delta)^{n_1}\gamma^{n-n_1} + (\gamma+\delta)^{n_2}\gamma^{n-n_2} + \chi_0\right), \tag{3.10}$$

$$w^\delta(\alpha_1 \wedge \alpha_2 \wedge \phi) = \frac{1}{2^p}\left((\gamma+\delta)^{n_1+1}\gamma^{n+1-n_1} + (\gamma+\delta)^{n_2+1}\gamma^{n+1-n_2} + \gamma^2\chi_0\right), \tag{3.11}$$

$$w^\delta(\alpha_2 \wedge \phi) = \frac{1}{2^p}\left((\gamma+\delta)^{n_1}\gamma^{n+1-n_1} + (\gamma+\delta)^{n_2+1}\gamma^{n-n_2} + \gamma\chi_0\right). \tag{3.12}$$

Substituting (3.9) to (3.12) into (3.8) and canceling the common factor of $\frac{1}{2^p}$ gives

$$\begin{aligned}&\frac{(\gamma+\delta)^{n_1+1}\gamma^{n-n_1} + (\gamma+\delta)^{n_2}\gamma^{n+1-n_2} + \gamma\chi_0}{(\gamma+\delta)^{n_1}\gamma^{n-n_1} + (\gamma+\delta)^{n_2}\gamma^{n-n_2} + \chi_0} \\ &\qquad \geq \frac{(\gamma+\delta)^{n_1+1}\gamma^{n+1-n_1} + (\gamma+\delta)^{n_2+1}\gamma^{n+1-n_2} + \gamma^2\chi_0}{(\gamma+\delta)^{n_1}\gamma^{n+1-n_1} + (\gamma+\delta)^{n_2+1}\gamma^{n-n_2} + \gamma\chi_0}.\end{aligned} \tag{3.13}$$

To simplify our notation, we take $x = \gamma + \delta$. We can now re-write (3.13) as

$$\frac{x^{n_1+1}\gamma^{n-n_1} + x^{n_2}\gamma^{n+1-n_2} + \gamma\chi_0}{x^{n_1}\gamma^{n-n_1} + x^{n_2}\gamma^{n-n_2} + \chi_0} \geq \frac{x^{n_1+1}\gamma^{n+1-n_1} + x^{n_2+1}\gamma^{n+1-n_2} + \gamma^2\chi_0}{x^{n_1}\gamma^{n+1-n_1} + x^{n_2+1}\gamma^{n-n_2} + \gamma\chi_0}.$$

[4]Indeed, when $\delta = 0$ we get w_I which we have also already proven, but also note that the forthcoming derivation is valid for this case.

Cross multiplying gives (noting that both denominators are non-negative)

$$\begin{aligned}
& x^{2n_1+1}\gamma^{2n-2n_1+1} + x^{n_1+n_2+2}\gamma^{2n-n_1-n_2} + x^{n_1+1}\gamma^{n-n_1+1}\chi_0 \\
& \quad + x^{n_1+n_2}\gamma^{2n-n_1-n_2+2} + x^{2n_2+1}\gamma^{2n-2n_2+1} + x^{n_2}\gamma^{n-n_2+2}\chi_0 \\
& \quad + x^{n_1}\gamma^{n-n_1+2}\chi_0 + x^{n_2+1}\gamma^{n-n_2+1}\chi_0 + \gamma^2\chi_0^2 \\
\geq\ & x^{2n_1+1}\gamma^{2n-2n_1+1} + x^{n_1+n_2+1}\gamma^{2n-n_1-n_2+1} + x^{n_1+1}\gamma^{n-n_1+1}\chi_0 \\
& \quad + x^{n_1+n_2+1}\gamma^{2n-n_1-n_2+1} + x^{2n_2+1}\gamma^{2n-2n_2+1} + x^{n_2+1}\gamma^{n-n_2+1}\chi_0 \\
& \quad + x^{n_1+1}\gamma^{n-n_1+2}\chi_0 + x^{n_2}\gamma^{n-n_2+2}\chi_0 + \gamma^2\chi_0^2
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
& x^{n_1+n_2+2}\gamma^{2n-n_1-n_2} + x^{n_1+n_2}\gamma^{2n-n_1-n_2+2} + x^{n_2}\gamma^{n-n_2+2}\chi_0 + x^{n_2+1}\gamma^{n-n_2+1}\chi_0 \\
\geq\ & x^{n_1+n_2+1}\gamma^{2n-n_1-n_2+1} + x^{n_1+n_2+1}\gamma^{2n-n_1-n_2+1} + x^{n_2+1}\gamma^{n-n_2+1}\chi_0 \\
& \quad + x^{n_2}\gamma^{n-n_2+2}\chi_0
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
& x^{n_1+n_2+2}\gamma^{2n-n_1-n_2} + x^{n_1+n_2}\gamma^{2n-n_1-n_2+2} - x^{n_1+n_2+1}\gamma^{2n-n_1-n_2+1} \\
& \quad - x^{n_1+n_2+1}\gamma^{2n-n_1-n_2+1} \\
\geq\ & x^{n_2+1}\gamma^{n-n_2+1}\chi_0 + x^{n_2}\gamma^{n-n_2+2}\chi_0 - x^{n_2}\gamma^{n-n_2+2}\chi_0 - x^{n_2+1}\gamma^{n-n_2+1}\chi_0
\end{aligned}$$

$$\Longleftrightarrow$$

$$x^{n_1+n_2}\gamma^{2n-n_1-n_2}(x^2 + \gamma^2 - x\gamma - x\gamma) \ \ \geq \ \ 0$$

$$\Longleftrightarrow$$

$$x^{n_1+n_2}\gamma^{2n-n_1-n_2}(x-\gamma)^2 \ \ \geq \ \ 0, \tag{3.14}$$

and this is trivially true, therefore the w^δ satisfy NPIR (and SPIR) for $\delta \in [0,1]$.

When $\delta = 0$, $\gamma = x$ so (3.14) holds with equality, however when $\delta \in (0,1)$ we have that

$$x^{n_1+n_2} > 0,\ \gamma^{2n-n_1-n_2} > 0, (x-\gamma)^2 = \delta^2 > 0,$$

and therefore

$$x^{n_1+n_2}\gamma^{2n-n_1-n_2}(x-\gamma)^2 > 0.$$

Which means that NPIR (& therefore SPIR) is satisfied strictly for $\delta \in (0,1)$, as required. $\square$

The "strictness" parts of the above theorems are illuminating since they say that apart from some undesirable, extreme, choices of[5] λ, δ we have, in the case of SPIR for example, that an additional observation of the state α will strictly *increase* our belief that the next state will be α compared with our belief if we'd observed an additional observation of some other state instead.

This result would appear to back up the desirability of our two new relevance principles, since they seem to say that something sensible occurs when we would expect it to.

3.4 SPIR and All Probability Functions Satisfying Ax

In this section we broaden our attention to all probability functions that satisfy Ax, as opposed just the two continua of the previous section. We ask whether NPIR & SPIR are satisfied by all such functions, and if not, where are they satisfied?

For this section we will be considering NPIR & SPIR expressed as follows

NPIR :

$$w(\alpha_1 \mid \langle n_1, n_2, n_3, \dots, n_{2^p}\rangle) \geq w(\alpha_1 \mid \langle n_1, n_2+1, n_3, \dots, n_{2^p}\rangle). \tag{3.15}$$

SPIR :

$$w(\alpha_1 \mid \langle n_1+1, n_2, n_3, \dots, n_{2^p}\rangle) \geq w(\alpha_1 \mid \langle n_1, n_2+1, n_3, \dots, n_{2^p}\rangle). \tag{3.16}$$

[5]i.e. given that $\delta = 0$ and $\lambda \to \infty$ give the noninductive Independent Solution, w_I and $\delta = 1$ and $\lambda = 0$ give Carnap's excessively selective $\mathfrak{m}_0$

Where the vectors in the definitions are vector representations. We will be particularly looking at this problem in reference to the number of predicates in our language, p, and the relationship between the number of observations of the state descriptions within those languages, particularly those involving n_1 and n_2 which are directly involved in the two principles being studied.

3.4.1 Where Our Principles Hold

We first look at the areas in which we know that NPIR & SPIR hold for all probability functions satisfying Ax.

Theorem 3.4.1. *NPIR and SPIR hold for all probability functions satisfying Ex (& Ax) which are defined over a language containing just one predicate (i.e. p=1).*

Proof. By Theorem 3.2.3 we only need to show this for NPIR. Since our language contains only one predicate there will be just two atoms namely $\alpha_1(x) = P_1(x)$ and $\alpha_2(x) = \neg P_1(x)$. Here our statement of NPIR given by (3.15) becomes

$$w(\alpha_1 \mid \langle n_1, n_2 \rangle) \geq w(\alpha_1 \mid \langle n_1, n_2 + 1 \rangle),$$

by de Finetti's Theorem this is equivalent to

$$\frac{\int x_1^{n_1+1} x_2^{n_2} d\mu}{\int x_1^{n_1} x_2^{n_2} d\mu} \geq \frac{\int x_1^{n_1+1} x_2^{n_2+1} d\mu}{\int x_1^{n_1} x_2^{n_2+1} d\mu}, \tag{3.17}$$

for some measure μ. Since, by definition we have that $\langle x_1, x_2 \rangle \in \mathbb{D}_2$ we must have that $x_2 = 1 - x_1$. Using this fact we get that (3.17) is equivalent to

$$\begin{aligned}
& \frac{\int x_1^{n_1+1} x_2^{n_2} d\mu}{\int x_1^{n_1} x_2^{n_2} d\mu} \quad \geq \quad \frac{\int (1-x_1) x_1^{n_1+1} x_2^{n_2} d\mu}{\int (1-x_1) x_1^{n_1} x_2^{n_2} d\mu} \\
\iff \; & \frac{\int x_1^{n_1+1} x_2^{n_2} d\mu}{\int x_1^{n_1} x_2^{n_2} d\mu} \quad \geq \quad \frac{\int x_1^{n_1+1} x_2^{n_2} d\mu - \int x_1^{n_1+2} x_2^{n_2} d\mu}{\int x_1^{n_1} x_2^{n_2} d\mu - \int x_1^{n_1+1} x_2^{n_2} d\mu}
\end{aligned}$$

$$\begin{aligned}
\iff \quad & \int x_1^{n_1+1} x_2^{n_2} d\mu \int x_1^{n_1} x_2^{n_2} d\mu - \int x_1^{n_1+1} x_2^{n_2} d\mu \int x_1^{n_1+1} x_2^{n_2} d\mu \\
& \geq \int x_1^{n_1} x_2^{n_2} d\mu \int x_1^{n_1+1} x_2^{n_2} d\mu - \int x_1^{n_1} x_2^{n_2} d\mu \int x_1^{n_1+2} x_2^{n_2} d\mu
\end{aligned}$$

$$\begin{array}{rcl}
\iff -\int x_1^{n_1+1}x_2^{n_2}d\mu \int x_1^{n_1+1}x_2^{n_2} & \geq & -\int x_1^{n_1}x_2^{n_2}d\mu \int x_1^{n_1+2}x_2^{n_2}d\mu \\
\iff \int x_1^{n_1+1}x_2^{n_2}d\mu \int x_1^{n_1+1}x_2^{n_2} & \leq & \int x_1^{n_1}x_2^{n_2}d\mu \int x_1^{n_1+2}x_2^{n_2}d\mu \\
& \iff & \\
\frac{\int x_1^{n_1+1}x_2^{n_2}d\mu}{\int x_1^{n_1}x^{n_2}d\mu} & \leq & \frac{\int x_1^{n_1+2}x_2^{n_2}d\mu}{\int x_1^{n_1+1}x_2^{n_2}d\mu} \\
& \iff & \\
w(\alpha_1 \mid \langle n_1, n_2\rangle) & \leq & w(\alpha_1 \mid \langle n_1+1, n_2\rangle),
\end{array}$$

and this is just an instance of PIR and therefore NPIR (& SPIR) hold in the case where $p = 1$, as required. □

If we now allow the number of predicates in our language to be picked arbitrarily then the situation becomes a bit murkier. We can, however, prove the following theorem.

Theorem 3.4.2. *NPIR & SPIR, given as defined in (3.15) and (3.16), are consequences of Ax in the following cases :*

1. $n_1 \geq n_2 = n_3 = n_4 = \ldots = n_{2^p}$,

2. $n_2 \geq n_1 = n_3 = n_4 = \ldots = n_{2^p}$.

Further, SPIR is a consequence of Ax in the two additional cases[6]*:*

3. $n_1 = n_2$,

4. $n_1 + 1 = n_2$.

Proof. We will take each case in turn. By Theorem 3.2.3 we only need to show cases 1 and 2 for NPIR.

__**Case One :**__ $n_1 \geq n_2 = n_3 = n_4 = \ldots = n_{2^p}$.

Here we may assume that $n_2 = n_3 = \ldots = n_{2^p} = 0$ for if not then we can replace the de Finetti prior μ by $(x_1x_2x_3\ldots x_{2^p})^{n_1}\mu$. Our statement of NPIR given by (3.15) therefore becomes

$$w(\alpha_1 \mid \langle n_1, 0, 0, \ldots, 0\rangle) \geq w(\alpha_1 \mid \langle n_1, 1, 0, \ldots, 0\rangle),$$

and by de Finetti's Theorem this is equivalent to

$$\frac{\int x_1^{n_1+1}d\mu}{\int x_1^{n_1}d\mu} \geq \frac{\int x_1^{n_1+1}x_2 d\mu}{\int x_1^{n_1}x_2 d\mu}, \tag{3.18}$$

[6] *See section 3.4.3 for what happens with NPIR here.*

for some measure μ. By Ax we have that

$$\int x_1^{n_1+1} x_2 d\mu = \frac{1}{2^p - 1} \int (1 - x_1) x_1^{n_1+1} d\mu,$$

and

$$\int x_1^{n_1} x_2 d\mu = \frac{1}{2^p - 1} \int (1 - x_1) x_1^{n_1} d\mu.$$

Substituting these in to (3.18) gives

$$\begin{aligned}
\frac{\int x_1^{n_1+1} d\mu}{\int x_1^{n_1} d\mu} &\geq \frac{\frac{1}{2^p-1} \int (1 - x_1) x_1^{n_1+1} d\mu}{\frac{1}{2^p-1} \int (1 - x_1) x_1^{n_1} d\mu} \\
\iff \frac{\int x_1^{n_1+1} d\mu}{\int x_1^{n_1} d\mu} &\geq \frac{\int x_1^{n_1+1} d\mu - \int x_1^{n_1+2} d\mu}{\int x_1^{n_1} d\mu - \int x_1^{n_1+1} d\mu}
\end{aligned}$$

$$\begin{aligned}
\iff \quad & \int x_1^{n_1+1} d\mu \int x^{n_1} d\mu - \int x_1^{n_1+1} d\mu \int x_1^{n_1+1} d\mu \\
& \geq \int x_1^{n_1} d\mu \int x_1^{n_1+1} d\mu - \int x_1^{n_1} d\mu \int x_1^{n_1+2} d\mu
\end{aligned}$$

$$\begin{aligned}
\iff -\int x_1^{n_1+1} d\mu \int x_1^{n_1+1} d\mu &\geq -\int x_1^{n_1} d\mu \int x_1^{n_1+2} d\mu \\
\iff \int x_1^{n_1+1} d\mu \int x_1^{n_1+1} d\mu &\leq \int x_1^{n_1} d\mu \int x_1^{n_1+2} d\mu \\
&\iff \\
\frac{\int x_1^{n_1+1} d\mu}{\int x_1^{n_1} d\mu} &\leq \frac{\int x_1^{n_1+2} d\mu}{\int x_1^{n_1+1} d\mu} \\
&\iff \\
w(\alpha_1 \mid \langle n_1, 0, 0, \ldots, 0 \rangle) &\leq w(\alpha_1 \mid \langle n_1 + 1, 0, 0, \ldots, 0 \rangle),
\end{aligned}$$

which is an instance of PIR, and so NPIR (& SPIR) both hold for this case, as required.

Case Two : $n_2 \geq n_1 = n_3 = n_4 = \ldots = n_{2^p}$.

Again we can take $n_1 = n_3 = \ldots = n_{2^p} = 0$, making our statement of NPIR here to be:

$$w(\alpha_1 \mid \langle 0, n_2, 0, \ldots, 0 \rangle) \geq w(\alpha_1 \mid \langle 0, n_2 + 1, 0, \ldots, 0 \rangle),$$

and by de Finetti's Theorem this is equivalent to

$$\frac{\int x_1 x_2^{n_2} d\mu}{\int x_2^{n_2} d\mu} \geq \frac{\int x_1 x_2^{n_2+1} d\mu}{\int x_2^{n_2+1} d\mu}, \tag{3.19}$$

for some measure μ. By Ax we have that

$$\int x_1 x_2^{n_2} d\mu = \frac{1}{2^p - 1} \int (1 - x_2) x_2^{n_2} d\mu,$$
$$\int x_1 x_2^{n_2+1} d\mu = \frac{1}{2^p - 1} \int (1 - x_2) x_2^{n_2+1} d\mu.$$

Substituting these into (3.19) gives

$$\begin{aligned}
\frac{\frac{1}{2^p-1} \int (1 - x_2) x_2^{n_2} d\mu}{\int x_2^{n_2} d\mu} &\geq \frac{\frac{1}{2^p-1} \int (1 - x_2) x_2^{n_2+1} d\mu}{\int x_2^{n_2+1} d\mu} \\
\Longleftrightarrow \quad \frac{\int x_2^{n_2} d\mu - \int x_2^{n_2+1} d\mu}{\int x_2^{n_2} d\mu} &\geq \frac{\int x_2^{n_2+1} d\mu - \int x_2^{n_2+2} d\mu}{\int x_2^{n_2+1} d\mu} \\
\Longleftrightarrow \quad \int x_2^{n_2} d\mu \int x_2^{n_2+1} d\mu &- \int x_2^{n_2+1} x_2^{n_2+1} d\mu \\
&\geq \int x_2^{n_2+1} d\mu \int x_2^{n_2} d\mu - \int x_2^{n_2+2} d\mu \int x_2^{n_2} \\
\Longleftrightarrow - \int x_2^{n_2+1} d\mu \int x_2^{n_2+1} d\mu &\geq - \int x_2^{n_2+2} d\mu \int x_2^{n_2} d\mu \\
\Longleftrightarrow \int x_2^{n_2+1} d\mu \int x_2^{n_2+1} d\mu &\leq \int x_2^{n_2+2} d\mu \int x_2^{n_2} d\mu \\
&\Longleftrightarrow \\
\frac{\int x_2^{n_2+1} d\mu}{\int x_2^{n_2} d\mu} &\leq \frac{\int x_2^{n_2+2} d\mu}{\int x_2^{n_2+1} d\mu} \\
&\Longleftrightarrow \\
w(\alpha_2 \mid \langle 0, n_2, 0, \ldots, 0 \rangle) &\leq w(\alpha_2 \mid \langle 0, n_2 + 1, 0, \ldots, 0 \rangle),
\end{aligned}$$

and again this is just an instance of PIR. Therefore case 2 holds for both NPIR and SPIR, as required.

Case Three : $n_1 = n_2$.

Recall that we will only prove this case for SPIR. Our statement of SPIR here, based upon that given in (3.16) is

$$w(\alpha_1 \mid \langle n_1 + 1, n_1, n_3, n_4, \ldots n_{2^p} \rangle) \geq w(\alpha_1 \mid \langle n_1, n_1 + 1, n_3, n_4, \ldots, n_{2^p} \rangle).$$

Expanding the conditional probabilities (and expressing the state descriptions involved in terms of their vector representations) gives:

$$\frac{w(\langle n_1+2, n_1, n_3, n_4, \ldots, n_{2^p}\rangle)}{w(\langle n_1+1, n_1, n_3, n_4, \ldots, n_{2^p}\rangle)} \geq \frac{w(\langle n_1+1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle)}{w(\langle n_1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle)}. \tag{3.20}$$

By Ax we have that

$$w(\langle n_1+1, n_1, n_3, n_4, \ldots, n_{2^p}\rangle) = w(\langle n_1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle),$$

so we can cancel the denominators in (3.20) to leave

$$w(\langle n_1+2, n_1, n_3, n_4, \ldots, n_{2^p}\rangle) \geq w(\langle n_1+1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle),$$

and this is an instance of The Only Rule, therefore SPIR holds in this case, as required.

Case Four : $n_1 + 1 = n_2$.

Again, we're only going to prove this case for SPIR. Our statement of SPIR here, based upon that given in (3.16) is

$$w(\alpha_1 \mid \langle n_1+1, n_1+1, n_3, n_4, \ldots n_{2^p}\rangle) \geq w(\alpha_1 \mid \langle n_1, n_1+2, n_3, n_4, \ldots, n_{2^p}\rangle).$$

Expanding the conditional probabilities gives :

$$\frac{w(\langle n_1+2, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle)}{w(\langle n_1+1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle)} \geq \frac{w(\langle n_1+1, n_1+2, n_3, n_4, \ldots, n_{2^p}\rangle)}{w(\langle n_1, n_1+2, n_3, n_4, \ldots, n_{2^p}\rangle)}. \tag{3.21}$$

By Ax we have that

$$w(\langle n_1+2, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle) = w(\langle n_1+1, n_1+2, n_3, n_4, \ldots, n_{2^p}\rangle),$$

so this time we can cancel the numerators in (3.21) to leave

$$\begin{array}{ccc} \dfrac{1}{w(\langle n_1+1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle)} & \geq & \dfrac{1}{w(\langle n_1, n_1+2, n_3, n_4, \ldots, n_{2^p}\rangle)} \\ & \iff & \\ w(\langle n_1, n_1+2, n_3, n_4, \ldots, n_{2^p}\rangle) & \geq & w(\langle n_1+1, n_1+1, n_3, n_4, \ldots, n_{2^p}\rangle) \end{array}$$

and this is an instance of The Only Rule. Therefore SPIR holds in this case, as required.

We have shown all four cases, as required. $\square$

3.4.2 Where Our Principles Do Not Hold & Some Unresolved Cases

After several attempts to prove that SPIR holds for all probability functions that satisfy Ax, we came to realise that this was, in fact, not the case. As such, we now move on to some symmetric probability functions which provide counter-examples to SPIR for certain choices of $n_1, n_2, \ldots, n_{2^p}$. Note that by Corollary 3.2.4 wherever we can find a counter-example to SPIR we can also find a counter-example to NPIR so, unless otherwise stated, all the following results in this section apply to both principles.

We first look at the case where $p \geq 3, n_1 \geq n_2 + 1 > 1$ and $n_3, \ldots, n_{2^p} = 0$.

For $\frac{1}{2^p} < \tau < \frac{1}{2}$, $\epsilon = \frac{1-\tau}{2^p-1}$ (note that this means that $\tau > \epsilon$) let[7]

$$\vec{a} = \langle \tau, \epsilon, \epsilon, \ldots, \epsilon \rangle \in \mathbb{D}_{2^p},$$

and,

$$\vec{b} = \langle \frac{1}{2}, \frac{1}{2}, 0, 0, \ldots, 0 \rangle \in \mathbb{D}_{2^p}.$$

Let w be the SM function based upon the measure μ that puts weight $\frac{2}{2+(2^p-1)c}$ on $\vec{a}$ and $\frac{(2^p-1)c}{2+(2^p-1)c}$ on $\vec{b}$ for very large c, then we have that

$$\begin{aligned} & w(\langle n_1 + 2, n_2, 0, \ldots, 0 \rangle) \\ &= \frac{1}{2^p!} \cdot \frac{1}{2 + (2^p - 1)c} \big(2((2^p - 1)!\tau^{n_1+2}\epsilon^{n_2} + (2^p - 1)!\epsilon^{n_1+2}\tau^{n_2} + \\ &\qquad (2^p - 2)(2^p - 1)!\epsilon^{n_1+n_2+2}) + (2^p - 1)c(2(2^p - 2)!2^{-(n_1+n_2+2)}) \big) \\ &= \frac{2}{2^p(2 + (2^p - 1)c)} \left(\tau^{n_1+2}\epsilon^{n_2} + \epsilon^{n_1+2}\tau^{n_2} + (2^p - 2)\epsilon^{n_1+n_2+2} + 2^{-(n_1+n_2+2)}c \right). \end{aligned}$$

Similarly,

$$\begin{aligned} & w(\langle n_1 + 1, n_2, 0, \ldots, 0 \rangle) \\ &= \frac{2}{2^p(2 + (2^p - 1)c)} \left(\tau^{n_1+1}\epsilon^{n_2} + \epsilon^{n_1+1}\tau^{n_2} + (2^p - 2)\epsilon^{n_1+n_2+1} + 2^{-(n_1+n_2+1)}c \right), \end{aligned}$$

[7] The $\vec{a}$ used in this counter-example, and in those that follow, should not be confused with the constants, a_1, a_2, $\ldots$, sometimes used in the definition of sentences.

$$
\begin{aligned}
& w(\langle n_1+1, n_2+1, 0, \ldots, 0\rangle) \\
& \quad = \frac{2}{2^p(2+(2^p-1)c)}\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+2}+2^{-(n_1+n_2+2)}c\right),
\end{aligned}
$$

$$
\begin{aligned}
& w(\langle n_1, n_2+1, 0, \ldots, 0\rangle) \\
& \quad = \frac{2}{2^p(2+(2^p-1)c)}\left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+1}+2^{-(n_1+n_2+1)}c\right),
\end{aligned}
$$

and so, by taking out the common factor of $\frac{2}{2^p(2+(2^p-1)c)}$ SPIR in this case becomes

$$
\begin{aligned}
& \frac{\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2}+(2^p-2)\epsilon^{n_1+n_2+2}+2^{-(n_1+n_2+2)}c}{\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p-2)\epsilon^{n_1+n_2+1}+2^{-(n_1+n_2+1)}c} \\
& \qquad \geq \frac{\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+2}+2^{-(n_1+n_2+2)}c}{\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+1}+2^{-(n_1+n_2+1)}c}.
\end{aligned}
$$

We now cross multiply. Since c can be made to be as large as we like, and the terms involving c^2 cancel, we only need to consider those terms that contain c, so if SPIR holds it must be the case that

$$
\begin{aligned}
& (2^{-(n_1+n_2+2)}c)\left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+1}\right) \\
& \qquad +(2^{-(n_1+n_2+1)}c)\left(\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2}+(2^p-2)\epsilon^{n_1+n_2+2}\right) \\
\geq\ & (2^{-(n_1+n_2+2)}c)\left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p-2)\epsilon^{n_1+n_2+1}\right) \\
& \qquad +(2^{-(n_1+n_2+1)}c)\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+2}\right).
\end{aligned}
$$

Taking out the common factor of $2^{-(n_1+n_2+1)}c$ leaves

$$
\begin{aligned}
& \left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+1}\right) \\
& \qquad +2\left(\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2}+(2^p-2)\epsilon^{n_1+n_2+2}\right) \\
& \quad \geq \left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p-2)\epsilon^{n_1+n_2+1}\right) \\
& \qquad\quad +2\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p-2)\epsilon^{n_1+n_2+2}\right)
\end{aligned}
$$

$$\Longleftrightarrow$$

$$
\begin{aligned}
& \tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+2\tau^{n_1+2}\epsilon^{n_2}+2\epsilon^{n_1+2}\tau^{n_2} \\
& \quad \geq \tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+2\tau^{n_1+1}\epsilon^{n_2+1}+2\epsilon^{n_1+1}\tau^{n_2+1}
\end{aligned} \tag{3.22}
$$

$$\Longleftrightarrow$$

$$
\begin{aligned}
&(\tau^{n_2}\epsilon^{n_2})(\tau^{n_1-n_2}\epsilon + \epsilon^{n_1-n_2}\tau + 2\tau^{n_1-n_2+2} + 2\epsilon^{n_1-n_2+2}) \\
&\qquad \geq (\tau^{n_2}\epsilon^{n_2})(\tau^{n_1-n_2+1} + \epsilon^{n_1-n_2+1} + 2\tau^{n_1-n_2+1}\epsilon + 2\epsilon^{n_1-n_2+1}\tau)
\end{aligned}
$$

$$\Longleftrightarrow$$

$$
\begin{aligned}
&\tau^{n_1-n_2}\epsilon + 2\tau^{n_1-n_2+2} - \tau^{n_1-n_2+1} - 2\tau^{n_1-n_2+1}\epsilon \\
&\qquad \geq \epsilon^{n_1-n_2+1} + 2\epsilon^{n_1-n_2+1}\tau - \epsilon^{n_1-n_2}\tau - 2\epsilon^{n_1-n_2+2}
\end{aligned}
$$

$$\Longleftrightarrow$$

$$
\begin{aligned}
\tau^{n_1-n_2}(\epsilon + 2\tau^2 - \tau - 2\tau\epsilon) &\geq \epsilon^{n_1-n_2}(\epsilon + 2\epsilon\tau - \tau - 2\epsilon^2) \\
\Longleftrightarrow \tau^{n_1-n_2}(1-2\tau)(\epsilon-\tau) &\geq \epsilon^{n_1-n_2}(1-2\epsilon)(\epsilon-\tau) \\
\Longleftrightarrow \tau^{n_1-n_2}(1-2\tau) &\leq \epsilon^{n_1-n_2}(1-2\epsilon) \\
\Longleftrightarrow \frac{\tau^{n_1-n_2}}{\epsilon^{n_1-n_2}} &\leq \frac{1-2\epsilon}{1-2\tau} \\
\Longleftrightarrow \left(\frac{\tau}{\epsilon}\right)^{n_1-n_2} &\leq \frac{1-2\epsilon}{1-2\tau}. \qquad (3.23)
\end{aligned}
$$

If we can find τ such that

$$\frac{\tau}{\epsilon} > \frac{1-2\epsilon}{1-2\tau}, \qquad (3.24)$$

then that τ gives a counter-example to SPIR for any value of $n_1 - n_2 \geq 1$, since $\frac{\tau}{\epsilon} > 1$ and the RHS of (3.23) is fixed as $n_1 - n_2$ increases. It turns out that setting $\tau = \frac{1}{4}$ suffices, as we will now show.

Setting $\tau = \frac{1}{4}$ gives $\epsilon = \frac{3}{4(2^p-1)}$ and therefore we have that (3.24) becomes

$$
\begin{aligned}
\frac{\frac{1}{4}}{\frac{3}{4(2^p-1)}} &> \frac{1-2\frac{3}{4(2^p-1)}}{\frac{1}{2}} \\
\Longleftrightarrow \frac{1}{8} &> \frac{3}{4(2^p-1)} - 2\left(\frac{3}{4(2^p-1)}\right)^2 \\
\Leftarrow \frac{1}{8} &> \frac{3}{4(2^p-1)} \\
\Leftarrow \frac{1}{8} &> \frac{3}{28},
\end{aligned}
$$

since $p \geq 3$, and this is trivially true.

Therefore we have found a counter-example to SPIR for any $p \geq 3$ and any $n_1 \geq n_2 + 1 > 1$ where $n_3 = n_4 = \ldots = n_{2^p} = 0$, as required.

By considering the inequality derived above we can take our analysis a bit further. Recall that if we can find a τ such that

$$\frac{\tau}{\epsilon} > \frac{1-2\epsilon}{1-2\tau},$$

then that τ forms a counter-example to SPIR for all $n_1 \geq n_2 + 1$, with $p \geq 3$. Our aim now is to find the range of values of τ which form such counter-examples.

We start by substituting $\epsilon = \frac{1-\tau}{2^p-1}$, giving

$$\begin{aligned} \frac{\tau}{\frac{1-\tau}{2^p-1}} &> \frac{1-2\left(\frac{1-\tau}{2^p-1}\right)}{1-2\tau} \\ \iff \tau - 2\tau^2 &> \frac{1-\tau}{2^p-1} - 2\left(\frac{1-\tau}{2^p-1}\right)^2. \end{aligned}$$

For ease of notation we now set $d = 2^p - 1$ to give

$$\begin{aligned} \tau - 2\tau^2 &> \frac{1-\tau}{d} - 2\left(\frac{1-\tau}{d}\right)^2 \\ \iff d^2\tau - 2d^2\tau^2 &> d - d\tau - 2(1-\tau)^2 \\ \iff d^2\tau - 2d^2\tau^2 &> d - d\tau - 2 + 4\tau - 2\tau^2 \\ \iff (2-2d^2)\tau^2 + (d^2+d-4)\tau + (2-d) &> 0. \end{aligned}$$

Since $d > 1$ this quadratic always forms a $\cap$-shaped parabola, and as such will be positive in the region bounded by the two real[8] solutions to

$$(2-2d^2)\tau^2 + (d^2+d-4)\tau + (2-d) = 0.$$

These two solutions are given by

$$\tau = \frac{-(d^2+d-4) \pm \sqrt{(d^2+d-4)^2 - 4(2-2d^2)(2-d)}}{2(2-2d^2)}. \tag{3.25}$$

[8]If there are not two real solutions then there will be no positive region for our graph. If this were the case it would then not be possible to give a τ which would form a counter-example.

Now

$$
\begin{aligned}
& (d^2 + d - 4)^2 - 4(2 - 2d^2)(2 - d) \\
= \; & (d^4 + d^3 + d^2 - 4d^2 + d^2 - 4d^2 - 4d - 4d + 16) - (16 - 8d - 16d^2 + 8d^3) \\
= \; & (d^4 + 2d^3 - 7d^2 - 8d + 16) - (16 - 8d - 16d^2 + 8d^3) \\
= \; & d^4 - 6d^3 + 9d^2 \\
= \; & d^2(d - 3)^2.
\end{aligned}
$$

So (3.25) becomes

$$
\begin{aligned}
\tau &= \frac{-(d^2 + d - 4) \pm \sqrt{d^2(d-3)^2}}{2(2 - 2d^2)} \\
\tau &= \frac{-d^2 - d + 4 \pm (d^2 - 3d)}{4 - 4d^2} \\
\tau &= \frac{-d^2 - d + 4 + d^2 - 3d}{(2+2d)(2-2d)} \quad \text{or} \quad \frac{-d^2 - d + 4 - d^2 + 3d}{(2+2d)(2-2d)} \\
\tau &= \frac{-4d + 4}{(2+2d)(2-2d)} \quad \text{or} \quad \frac{-2d^2 + 2d + 4}{(2+2d)(2-2d)} \\
\tau &= \frac{2(2 - 2d)}{(2+2d)(2-2d)} \quad \text{or} \quad \frac{(2-d)(2+2d)}{(2+2d)(2-2d)} \\
\tau &= \frac{2}{2+2d} \quad \text{or} \quad \frac{2-d}{2-2d} \left(= \frac{d-2}{2d-2} \right).
\end{aligned}
$$

Note that

$$\frac{2}{2+2d} = \frac{2}{2 + 2^{p+1} - 2} = \frac{1}{2^p},$$

and[9]

$$\frac{d-2}{2d-2} < \frac{1}{2},$$

so the positive region of the graph, and therefore the region which gives counter-examples to SPIR of this form, can be found in the range $\tau \in \left(\frac{1}{2^p}, \frac{d-2}{2d-2}\right)$.

This analysis can be used to show that this type of counter-example can not be formed for the case $p = 2$ when $n_1 = n_2 + 1$ since then the range of values of τ which

[9] In fact, as $p \to \infty$, $\frac{d-2}{2d-2} \to \frac{1}{2}$.

provide counter-examples collapses to $\left(\frac{1}{4}, \frac{1}{4}\right)$, i.e. an empty region[10].

However, our prior analysis showed that this region widens as we increase the difference between n_1 and n_2, so what happens if we preform a similar analysis in the case $n_1 = n_2 + 2$, $p = 2$?

i.e. We are asking what range of τ satisfies the inequality

$$\frac{\tau^2}{\epsilon^2} > \frac{1-2\epsilon}{1-2\tau},$$

when $p = 2$.

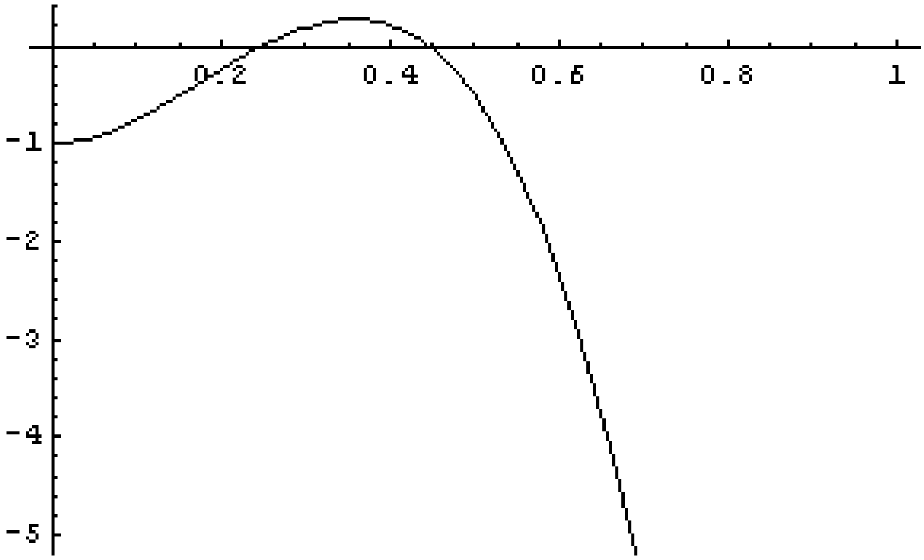
In this case $\epsilon = \frac{1-\tau}{3}$, substituting this in gives

$$\begin{aligned}
\frac{\tau^2}{\left(\frac{1-\tau}{3}\right)^2} &> \frac{1-2\left(\frac{1-\tau}{3}\right)}{1-2\tau} \\
\iff \tau^2 - 2\tau^3 &> \left(\frac{1-\tau}{3}\right)^2 - 2\left(\frac{1-\tau}{3}\right)^3 \\
\iff 27\tau^2 - 54\tau^3 &> 3(1-\tau)^2 - 2(1-\tau)^3 \\
\iff 27\tau^2 - 54\tau^3 &> 1 - 3\tau^2 + 2\tau^3 \\
\iff -1 + 30\tau^2 - 56\tau^3 &> 0.
\end{aligned}$$

The graph of this cubic[11] looks like

[10]It is worth pointing out, as an aside, that if we had adopted the Carnapian approach - which allows predicates representing mutually exclusive properties to be treated as atoms (a real life example would be the possible colours of an object) - then the same derivation shows that we can find counter-examples whenever $2^p > 4$, i.e. we can reduce the result given from 8 “colours” to 5. This footnote also applies to the derivation when we consider the forthcoming case $0 < n_1 \leq n_2 - 2, n_3 = n_4 = \ldots = n_{2^p} = 0$.

[11]Provided by Mathematica.



So the positive region will be found between the two positive roots of

$$-1 + 30\tau^2 - 56\tau^3 = 0.$$

Rather than go through the lengthy derivation of the roots, we use Mathematica (which uses Cardano's method) to find that the three solutions are

$$\tau = \frac{\left(2 - 3\sqrt{2}\right)}{14},\ \frac{\left(2 + 3\sqrt{2}\right)}{14},\ \frac{1}{4}.$$

So we can find a range where τ forms a counter-example for this case, namely $\left(\frac{1}{4}, \frac{\left(2+3\sqrt{2}\right)}{14}\right)$. A specific example is $\tau = \frac{2}{5}$ which gives $\vec{a} = \left\langle \frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right\rangle$.

We now move on to the case $0 < n_1 \leq n_2 - 2, n_3 = n_4 = \ldots = n_{2^p} = 0,\ p \geq 3$. Here our original τ-type set up will not work since as we decrease $n_1 - n_2$ our range of values (of τ) for which a counter-example can be found can only narrow (because the LHS of (3.23) will decrease whilst the RHS stays constant) and Theorem 3.4.2 gives that we cannot find a counter-example for the case $n_1 = n_2$. We therefore need to alter our approach.

For $\tau > \frac{1}{2}$, $\epsilon = 1 - \tau$ (i.e. $\tau > \epsilon$), let

$$\begin{aligned} \vec{a} &= \langle \tau, \epsilon, 0, \ldots, 0 \rangle \in \mathbb{D}_{2^p}, \\ \vec{b} &= \left\langle \frac{1}{2^p}, \frac{1}{2^p}, \ldots, \frac{1}{2^p} \right\rangle \in \mathbb{D}_{2^p}. \end{aligned}$$

Let w be the SM-function based upon the measure that puts weight $\frac{2^p(2^p-1)}{(2^p(2^p-1))+c}$ on the point $\vec{a}$ and $\frac{c}{(2^p(2^p-1))+c}$ on $\vec{b}$ for very large c, then we have that

$$\begin{aligned}
&w(\langle n_1+2, n_2, 0, \ldots, 0\rangle) \\
&\quad = \frac{1}{2^p!} \cdot \frac{1}{(2^p(2^p-1))+c} \\
&\qquad \left(2^p(2^p-1)\left((2^p-2)!(\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2})\right)+2^p!c(2^p)^{-(n_1+n_2+2)}\right) \\
&\quad = \frac{1}{(2^p(2^p-1))+c}\left(\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2}+(2^p)^{-(n_1+n_2+2)}c\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&w(\langle n_1+1, n_2, 0, \ldots, 0\rangle) \\
&\quad = \frac{1}{(2^p(2^p-1))+c}\left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p)^{-(n_1+n_2+1)}c\right),
\end{aligned}$$

$$\begin{aligned}
&w(\langle n_1+1, n_2+1, 0, \ldots, 0\rangle) \\
&\quad = \frac{1}{(2^p(2^p-1))+c}\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+2)}c\right),
\end{aligned}$$

$$\begin{aligned}
&w(\langle n_1, n_2+1, 0, \ldots, 0\rangle) \\
&\quad = \frac{1}{(2^p(2^p-1))+c}\left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+1)}c\right).
\end{aligned}$$

Removing the common factor of $\frac{1}{(2^p(2^p-1))+c}$ gives that SPIR in this case is equivalent to;

$$\begin{aligned}
&\frac{\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2}+(2^p)^{-(n_1+n_2+2)}c}{\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p)^{-(n_1+n_2+1)}c} \\
&\qquad \geq \frac{\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+2)}c}{\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+1)}c}
\end{aligned}$$

We now cross multiply. As before we only need to consider the terms that involve c, which leaves

$$\begin{aligned}
&(2^p)^{-(n_1+n_2+1)}c\left(\tau^{n_1+2}\epsilon^{n_2}+\epsilon^{n_1+2}\tau^{n_2}\right)+(2^p)^{-(n_1+n_2+2)}c\left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}\right) \\
&\quad \geq (2^p)^{-(n_1+n_2+1)}c\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}\right) \\
&\qquad\qquad + (2^p)^{-(n_1+n_2+2)}c\left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}\right).
\end{aligned}$$

We now remove the common factor of $(2^p)^{-(n_1+n_2+2)}\tau^{n_1}\epsilon^{n_1}c$ to give

$$\begin{aligned} & 2^p(\tau^2\epsilon^{n_2-n_1} + \epsilon^2\tau^{n_2-n_1}) + \epsilon^{n_2-n_1+1} + \tau^{n_2-n_1+1} \\ & \quad \geq 2^p(\tau\epsilon^{n_2-n_1+1} + \epsilon\tau^{n_2-n_1+1}) + \tau\epsilon^{n_2-n_1} + \epsilon\tau^{n_2-n_1} \end{aligned}$$

$$\iff$$

$$\begin{aligned} & 2^p\tau^{n_2-n_1}\epsilon^2 + \tau^{n_2-n_1+1} - 2^p\tau^{n_2-n_1+1}\epsilon - \tau^{n_2-n_1}\epsilon \\ & \quad \geq 2^p\epsilon^{n_2-n_1+1}\tau + \epsilon^{n_2-n_1}\tau - 2^p\epsilon^{n_2-n_1}\tau^2 - \epsilon^{n_2-n_1+1} \end{aligned}$$

$$\iff$$

$$\begin{aligned} \tau^{n_2-n_1}(\epsilon-\tau)(2^p\epsilon-1) &\geq \epsilon^{n_2-n_1}(\epsilon-\tau)(2^p\tau-1) \\ \tau^{n_2-n_1}(2^p\epsilon-1) &\leq \epsilon^{n_2-n_1}(2^p\tau-1). \end{aligned} \tag{3.26}$$

If $\epsilon \leq \frac{1}{2^p}$ then the LHS of (3.26) is ≤ 0, whilst the RHS will always be positive since $(2^p\tau - 1) > 0$, so we will not be able to find a counter-example to SPIR in that case. Therefore we restrict our attention to when $\frac{1}{2^p} < \epsilon < \frac{1}{2}$ (i.e. $\frac{1}{2} < \tau < \frac{2^p-1}{2^p}$), which allows us to derive that (3.26) is equivalent to:

$$\left(\frac{\tau}{\epsilon}\right)^{n_2-n_1} \leq \frac{2^p\tau-1}{2^p\epsilon-1},$$

so to find a counter-example we need to find $\tau \in \left(\frac{1}{2}, \frac{2^p-1}{2^p}\right)$ such that

$$\left(\frac{\tau}{\epsilon}\right)^{n_2-n_1} > \frac{2^p\tau-1}{2^p\epsilon-1}. \tag{3.27}$$

If we take $n_2 - n_1 = 1$ (which we already know satisfies SPIR due to Theorem 3.4.2) and substitute $\epsilon = 1 - \tau$ then we end up with (3.27) being equivalent to

$$2\tau < 1$$

which is impossible, backing up our earlier Theorem. Here we are interested in what happens when $n_2 - n_1 > 1$. Note that if we can find a range for τ when $n_2 - n_1 = 2$ then that range will hold for all values of $n_2 - n_1 \geq 2$ as the LHS of (3.27) will keep increasing whilst the RHS will stay constant. We therefore look directly at this case and work from there.

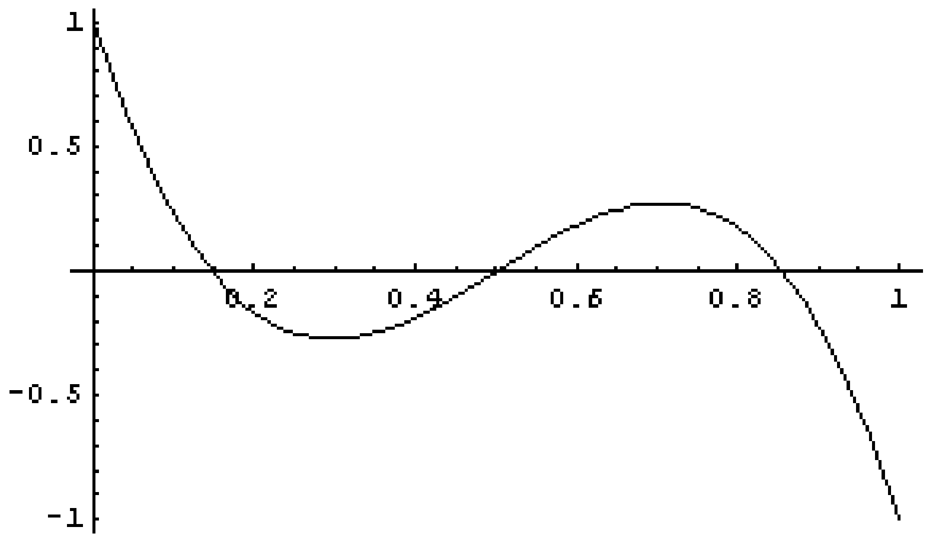
With $n_2 - n_1 = 2$, and recalling that $\epsilon = 1 - \tau$, we get that (3.27) is equivalent to

$$\begin{aligned}
\left(\frac{\tau}{\epsilon}\right)^2 &> \frac{2^p\tau - 1}{2^p\epsilon - 1} \\
\iff \tau^2(2^p(1-\tau) - 1) &> (1-\tau)^2(2^p\tau - 1) \\
\iff (2^p - 1)\tau^2 - 2^p\tau^3 &> -1 + (2^p + 2)\tau - (2\cdot 2^p + 1)\tau^2 + 2^p\tau^3 \\
\iff 1 - (2^p + 2)\tau + 3\cdot 2^p\tau^2 - 2\cdot 2^p\tau^3 &> 0.
\end{aligned}$$

The roots for the equation $1 - (2^p + 2)\tau + 3\cdot 2^p\tau^2 - 2\cdot 2^p\tau^3 = 0$ are[12]

$$\tau = \frac{\pm\sqrt{-4 + 2^p}\sqrt{2^p} + 2^p}{2\cdot 2^p},\ \tau = \frac{1}{2},$$

and the graph looks like[13]



So we can find counter-examples when τ is in the range $\left(\frac{1}{2}, \frac{1}{2} + \frac{\sqrt{-4+2^p}\sqrt{2^p}}{2\cdot 2^p}\right)$. Notice that the range not only increases as we increase $n_2 - n_1$ but it also increases as we increase p, in fact as we have that

$$\frac{\sqrt{-4 + 2^p}\sqrt{2^p}}{2\cdot 2^p} \longrightarrow_{p\to\infty} \frac{1}{2},$$

[12]From Mathematica.

[13]From Mathematica, this is specifically the graph when $p = 3$.

the range tends towards $\left(\frac{1}{2}, 1\right)$ (since $\frac{2^p-1}{2^p}$ also tends towards 1) i.e. the largest possible range we could have initially suspected.

Once again we start having problems when $p = 2$. Here our range collapses to $\left(\frac{1}{2}, \frac{1}{2}\right)$ and it turns out that this is our other main unresolved case - which as yet we have not been able to prove or find a counter-example for. We can, however, find a counter-example if we increase $n_2 - n_1$ to 3 (and, therefore, for any value that is greater than 3).

In this case (3.27) becomes

$$\begin{aligned} \frac{\tau^3}{(1-\tau)^3} &> \frac{4\tau - 1}{4(1-\tau)-1} \\ \iff \tau^3(3-4\tau) &> (4\tau-1)(1-\tau)^3 \\ \iff 1 - 7\tau + 15\tau^2 - 10\tau^3 &> 0. \end{aligned}$$

The roots of $1 - 7\tau + 15\tau^2 - 10\tau^3 = 0$ are

$$\tau = \frac{1}{10}\left(5 \pm \sqrt{5}\right), \ \frac{1}{2},$$

and since the graph looks similar to that on page 79, we have that we can find counter-examples whenever $\tau \in \left(\frac{1}{2}, \frac{\left(5+\sqrt{5}\right)}{10}\right)$.

All the counter-examples so far have been restricted to having $n_3 = n_4 = \ldots = n_{2^p} = 0$ so we provide an example to show that this doesn't necessarily need to be the case.

Here we are considering the version of SPIR defined by

$$\frac{w\left(\langle n_1+2, n_2, n_3, \ldots, n_{2^p}\rangle\right)}{w\left(\langle n_1+1, n_2, n_3, \ldots, n_{2^p}\rangle\right)} \geq \frac{w\left(\langle n_1+1, n_2+1, n_3, \ldots, n_{2^p}\rangle\right)}{w\left(\langle n_1, n_2+1, n_3, \ldots, n_{2^p}\rangle\right)}.$$

For $V > \frac{1}{2}$, $V > \tau \gg \epsilon = \frac{1-V-\tau}{2^p-2}$ let

$$\vec{a} = \langle V, \tau, \epsilon, \epsilon, \ldots, \epsilon\rangle \in \mathbb{D}_{2^p},$$

$$\vec{b} = \left\langle \frac{1}{2^p}, \frac{1}{2^p}, \ldots, \frac{1}{2^p} \right\rangle \in \mathbb{D}_{2^p}.$$

Let w be the SM-function based upon the measure which puts weight $\frac{1}{1+d}$ on $\vec{a}$ and $\frac{d}{1+d}$ on $\vec{b}$. Let

$$\begin{aligned}
x_1 &= (2^p - 2)! \sum_{i=2}^{2^p} \tau^{n_i} \left(\prod_{\substack{j \neq i \\ 2 \leq j \leq 2^p}} \epsilon^{n_j} \right), \\
x_2 &= (2^p - 2)! \sum_{i=2}^{2^p} V^{n_i} \left(\prod_{\substack{j \neq i \\ 2 \leq j \leq 2^p}} \epsilon^{n_j} \right), \\
x_3 &= (2^p - 3)! \sum_{\substack{i \neq j \\ 2 \leq i,j \leq 2^p}} V^{n_i} \tau^{n_j} \left(\prod_{\substack{k \neq i,j \\ 2 \leq k \leq 2^p}} \epsilon^{n_k} \right).
\end{aligned}$$

Let x'_1, x'_2, x'_3 be similarly defined but with n_2 replaced by $n_2 + 1$, then set

$$\begin{aligned}
X_1 &= V^{n_1+2} x_1 + \tau^{n_1+2} x_2 + (2^p - 2)\epsilon^{n_1+2} x_3, \\
X_2 &= V^{n_1+1} x_1 + \tau^{n_1+1} x_2 + (2^p - 2)\epsilon^{n_1+1} x_3, \\
X_3 &= V^{n_1+1} x'_1 + \tau^{n_1+1} x'_2 + (2^p - 2)\epsilon^{n_1+1} x'_3, \\
X_4 &= V^{n_1} x'_1 + \tau^{n_1} x'_2 + (2^p - 2)\epsilon^{n_1} x'_3, \\
Y &= 2^p! \left(\frac{1}{2^p} \right)^{N+1},
\end{aligned}$$

where $N = \sum_i n_i$. Then we have that

$$\begin{aligned}
2^p!(d+1) w(\langle n_1 + 2, n_2, n_3, \ldots, n_{2^p} \rangle) &= X_1 + \frac{dY}{2^p}, \\
2^p!(d+1) w(\langle n_1 + 1, n_2, n_3, \ldots, n_{2^p} \rangle) &= X_2 + dY, \\
2^p!(d+1) w(\langle n_1 + 1, n_2 + 1, n_3, \ldots, n_{2^p} \rangle) &= X_3 + \frac{dY}{2^p}, \\
2^p!(d+1) w(\langle n_1, n_2 + 1, n_3, \ldots, n_{2^p} \rangle) &= X_4 + dY.
\end{aligned}$$

So this instance of SPIR can be written as

$$\frac{X_1 + \frac{dY}{2^p}}{X_2 + dY} \geq \frac{X_3 + \frac{dY}{2^p}}{X_4 + dY}, \tag{3.28}$$

setting

$$d = \frac{2^p c}{Y},$$

gives that (3.28) is equivalent to

$$\frac{X_1 + c}{X_2 + 2^p c} \geq \frac{X_3 + c}{X_4 + 2^p c}.$$

Since c can be made as large as we like, and the c^2 terms cancel, we get that we get a counter-example to SPIR in this case if

$$2^p X_1 + X_4 \geq 2^p X_3 + X_2. \tag{3.29}$$

We are not going to continue the analysis past this point as it is overly complex whilst not shedding any new light on our unresolved cases. However once we have SPIR in the form given by (3.29) it becomes easier to state[14] a counter-example where the n_i, $i \geq 3$ are not all zero, as will now follow.

Take[15]

$$\vec{n} = \langle 17, 21, 3, 2, 2, 2, 1, 1 \rangle$$
$$V = \frac{55}{100}, \ \tau = \frac{44}{100}, \ \epsilon = \frac{1}{600}.$$

Then we have that[16] (all to 6 significant figures[17]):

$$\begin{aligned} X_1 &= 1.93077 \times 10^{-40}, \\ X_2 &= 4.04562 \times 10^{-40}, \\ X_3 &= 2.07440 \times 10^{-40}, \\ X_4 &= 4.44055 \times 10^{-40}, \end{aligned}$$

$$\begin{aligned} 8X_1 + X_4 &= 1.98867 \times 10^{-39}, \\ 8X_3 + X_2 &= 2.06408 \times 10^{-39}, \end{aligned}$$

and so we have

$$2^p X_1 + X_4 < 2^p X_3 + X_2,$$

[14]Indeed, it became easier to find a counter-example!

[15]It is worth pointing out that we are yet to find a counter-example where $\sum_{i=3}^{2^p} n_i \geq MAX(n_1, n_2)$.

[16]Excel was used to find the counter-example, and Mathematica to check it. The programme used for checking is given in Appendix A.

[17]Exact figures were used in all calculations.

which contradicts (3.29), as required.

In summary then, using the original τ-type set ups, with $n_3 = n_4 = \ldots = n_{2^p} = 0$, we can find counter examples to the cases

$$p \geq 3,\ n_1 \geq n_2 + 1,$$

minimally with τ in the range $\left(\frac{1}{2^p}, \frac{2^p-3}{2^{p+1}-4}\right)$, and

$$p = 2,\ n_1 \geq n_2 + 2,$$

minimally with τ in the range $\left(\frac{1}{4}, \frac{\left(2+3\sqrt{2}\right)}{14}\right)$, whilst the case

$$p = 2,\ n_1 = n_2 + 1, \tag{3.30}$$

remains unresolved. Additionally, when $n_3 = n_4 = \ldots = n_{2^p} = 0$, and using the second τ-type set up we can find counter examples to the cases

$$p \geq 3,\ n_1 \leq n_2 - 2$$

minimally with τ in the range $\left(\frac{1}{2}, \frac{1}{2} + \frac{\sqrt{-4+2^p}\sqrt{2^p}}{2 \cdot 2^p}\right)$, and

$$p = 2,\ n_1 \leq n_2 - 3$$

minimally with τ in the range $\left(\frac{1}{2}, \frac{5+\sqrt{5}}{10}\right)$, whilst the case

$$p = 2,\ n_1 = n_2 - 2 \tag{3.31}$$

remains unresolved. We also know that we need not restrict ourselves to $n_3 = n_4 = \ldots = n_{2^p} = 0$ when we are searching for counter-examples.

These unresolved cases have so far remained impervious to our attempts to find any type of counter-example to, which leads us to believe that it may well be the case that SPIR holds in those situations. However it has also remained impervious to our attempts to prove them too! We can improve the situation when we consider these unresolved cases in relation to NPIR as we will do in the next section.

3.4.3 NPIR and the Unresolved Cases

In total we have four situations where NPIR is as yet unresolved. These are cases three and four from Theorem 3.4.2 (where SPIR holds) and the those given above in (3.30) and (3.31) (where SPIR is unresolved). Namely these four cases are where $n_i = 0$ for $i \geq 3$ and

1. $n_1 = n_2$,
2. $n_1 + 1 = n_2$,
3. $p = 2$, $n_1 = n_2 + 1$,
4. $p = 2$, $n_1 + 2 = n_2$.

We will now show that we can find counter-examples to NPIR holding in all four of these cases.

Consider the following instance of NPIR;

$$\frac{w(\langle n_1+1, n_2, 0, \ldots, 0\rangle)}{w(\langle n_1, n_2, 0, \ldots, 0\rangle)} \geq \frac{w(\langle n_1+1, n_2+1, 0, \ldots, 0\rangle)}{w(\langle n_1, n_2+1, 0, \ldots, 0\rangle)}. \tag{3.32}$$

We are going to use our second τ-type set-up from the previous section, so for $\tau > \frac{1}{2}$, $\epsilon = 1 - \tau$, let

$$\begin{aligned} \vec{a} &= \langle \tau, \epsilon, 0, \ldots, 0\rangle \in \mathbb{D}_{2^p}, \\ \vec{b} &= \left\langle \frac{1}{2^p}, \frac{1}{2^p}, \ldots, \frac{1}{2^p} \right\rangle \in \mathbb{D}_{2^p}. \end{aligned}$$

Let w be the SM-function based upon the measure that puts weight $\frac{2^p(2^p-1)}{(2^p(2^p-1))+c}$ on the point $\vec{a}$ and $\frac{c}{(2^p(2^p-1))+c}$ on $\vec{b}$ for very large c. Then we have that

$$\begin{aligned} & w(\langle n_1+1, n_2, 0, \ldots, 0\rangle) \\ & \quad = \frac{1}{(2^p(2^p-1))+c}\left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p)^{-(n_1+n_2+1)}c\right), \end{aligned}$$

$$\begin{aligned} & w(\langle n_1, n_2, 0, \ldots, 0\rangle) \\ & \quad = \frac{1}{(2^p(2^p-1))+c}\left(\tau^{n_1}\epsilon^{n_2}+\epsilon^{n_1}\tau^{n_2}+(2^p)^{-(n_1+n_2)}c\right), \end{aligned}$$

$$\begin{aligned} & w(\langle n_1+1, n_2+1, 0, \ldots, 0\rangle) \\ & \quad = \frac{1}{(2^p(2^p-1))+c)}\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+2)}c\right), \end{aligned}$$

$$\begin{aligned} & w(\langle n_1, n_2+1, 0, \ldots, 0\rangle) \\ & \quad = \frac{1}{(2^p(2^p-1))+c)}\left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+1)}c\right), \end{aligned}$$

and so, by taking out the common factor of $\frac{1}{(2^p(2^p-1))+c)}$ we have that NPIR in this case is equivalent to

$$\begin{aligned} & \frac{\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+(2^p)^{-(n_1+n_2+1)}c}{\tau^{n_1}\epsilon^{n_2}+\epsilon^{n_1}\tau^{n_2}+(2^p)^{-(n_1+n_2)}c} \\ & \qquad \geq \frac{\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+2)}c}{\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}+(2^p)^{-(n_1+n_2+1)}c}. \end{aligned}$$

Again, we can make c as large as we like, so cross multiplying gives that we can find a counter-example to NPIR if we can find a counter-example to

$$\begin{aligned} & (2^p)^{-(n_1+n_2+1)}c\left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}\right)+(2^p)^{-(n_1+n_2+1)}c\left(\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}\right) \\ & \quad \geq (2^p)^{-(n_1+n_2)}c\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}\right)+(2^p)^{-(n_1+n_2+2)}c\left(\tau^{n_1}\epsilon^{n_2}+\epsilon^{n_1}\tau^{n_2}\right) \end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned} & 2^p\left(\tau^{n_1+1}\epsilon^{n_2}+\epsilon^{n_1+1}\tau^{n_2}+\tau^{n_1}\epsilon^{n_2+1}+\epsilon^{n_1}\tau^{n_2+1}\right) \\ & \quad \geq (2^p)^2\left(\tau^{n_1+1}\epsilon^{n_2+1}+\epsilon^{n_1+1}\tau^{n_2+1}\right)+\left(\tau^{n_1}\epsilon^{n_2}+\epsilon^{n_1}\tau^{n_2}\right). \end{aligned} \tag{3.33}$$

We will now consider each case in turn. For cases one and two we will substitute $d = 2^p$, for clarity (for cases three and four this won't be necessary as $2^p = 4$).

Case One : $n_1 = n_2$.

In this case (3.33) becomes

$$\begin{aligned} & d\left(\tau^{n_1+1}\epsilon^{n_1} + \epsilon^{n_1+1}\tau^{n_1} + \tau^{n_1}\epsilon^{n_1+1} + \epsilon^{n_1}\tau^{n_1+1}\right) \\ & \quad \geq d^2\left(\tau^{n_1+1}\epsilon^{n_1+1} + \epsilon^{n_1+1}\tau^{n_1+1}\right) + \left(\tau^{n_1}\epsilon^{n_1} + \epsilon^{n_1}\tau^{n_1}\right). \end{aligned}$$

If we take out the common factor of $\tau^{n_1}\epsilon^{n_1}$ then we get that this is equivalent to

$$d\left(\tau + \epsilon + \epsilon + \tau\right) \;\; \geq \;\; d^2\left(2\tau\epsilon\right) + 2,$$

substituting in $\epsilon = 1 - \tau$ gives that this is equivalent to

$$\begin{aligned} 2d & \geq d^2\left(2\tau(1-\tau)\right) + 2 \\ \iff 0 & \geq d^2\left(2\tau - 2\tau^2\right) - 2d + 2 \\ \iff 0 & \geq -d^2\tau^2 + d^2\tau + 1 - d. \end{aligned} \tag{3.34}$$

The roots of the quadratic

$$0 = -d^2\tau^2 + d^2\tau + 1 - d$$

are

$$\tau = \frac{d-1}{d}, \; \frac{1}{d},$$

and since the graph of the quadratic has a $\cap$ shape, we can find counter-examples to NPIR wherever $\tau \in \left(\frac{1}{2}, \frac{2^p-1}{2^p}\right)$, so this unresolved case for NPIR is cleared up. Note that the roots are symmetric around $\frac{1}{2}$, given how this probability function is set up and that $\tau = 1 - \epsilon$ this is entirely as you would expect.

Case Two : $n_1 + 1 = n_2$.

In this case (3.33) becomes

$$\begin{aligned} & d\left(\tau^{n_1+1}\epsilon^{n_1+1} + \epsilon^{n_1+1}\tau^{n_1+1} + \tau^{n_1}\epsilon^{n_1+2} + \epsilon^{n_1}\tau^{n_1+2}\right) \\ & \quad \geq d^2\left(\tau^{n_1+1}\epsilon^{n_1+2} + \epsilon^{n_1+1}\tau^{n_1+2}\right) + \left(\tau^{n_1}\epsilon^{n_1+1} + \epsilon^{n_1}\tau^{n_1+1}\right). \end{aligned}$$

If we take out the common factor of $\tau^{n_1}\epsilon^{n_1}$ then we get that this is equivalent to

$$d\left(\tau\epsilon + \epsilon\tau + \epsilon^2 + \tau^2\right) \geq d^2\left(\tau\epsilon^2 + \epsilon\tau^2\right) + (\epsilon + \tau)\,. \tag{3.35}$$

We now substitute $\epsilon = 1 - \tau$ in the above to give

$$\begin{aligned} d\left(2\tau(1-\tau) + (1-\tau)^2 + \tau^2\right) &\geq d^2\left(\tau(1-\tau)^2 + (1-\tau)\tau^2\right) + 1 \\ \iff d\left(2\tau - 2\tau^2 + 1 - 2\tau + \tau^2 + \tau^2\right) &\geq d^2\left(\tau - 2\tau^2 + \tau^3 + \tau^2 - \tau^3\right) + 1 \\ \iff d &\geq d^2\left(\tau - \tau^2\right) + 1 \\ \iff 0 &\geq -d^2\tau^2 + d^2\tau + 1 - d. \end{aligned}$$

Which is the same quadratic inequality as found in case one, (3.34). Similarly then, we can find counter-examples here whenever $\tau \in \left(\frac{1}{2}, \frac{d-1}{d}\right)$ and this unresolved case is cleared up.

Case Three : $p = 2, n_1 = n_2 + 1$.

In this case (3.33) becomes

$$\begin{aligned} &4\left(\tau^{n_2+2}\epsilon^{n_2} + \epsilon^{n_2+2}\tau^{n_2} + \tau^{n_2+1}\epsilon^{n_2+1} + \epsilon^{n_2+1}\tau^{n_2+1}\right) \\ &\quad \geq 16\left(\tau^{n_2+2}\epsilon^{n_2+1} + \epsilon^{n_2+2}\tau^{n_2+1}\right) + \left(\tau^{n_2+1}\epsilon^{n_2} + \epsilon^{n_2+1}\tau^{n_2}\right). \end{aligned}$$

If we take out the common factor of $\tau^{n_2}\epsilon^{n_2}$ then we get that this is equivalent to

$$4\left(\tau^2 + \epsilon^2 + 2\tau\epsilon\right) \geq 16\left(\tau^2\epsilon + \epsilon^2\tau\right) + \tau + \epsilon. \tag{3.36}$$

This is very similar to (3.35), with $d = 4$, and as such we end up with (3.36) being equivalent to the quadratic inequality

$$0 \geq -16\tau^2 + 16\tau - 3$$

and therefore, by following the same reasoning as for cases one and two, we can find counter-examples here when τ is in the range $\left(\frac{1}{2}, \frac{3}{4}\right)$. Hence this unresolved case is cleared up.

Case Four : $p = 2, n_1 + 2 = n_2$.

In this case (3.33) becomes

$$4\left(\tau^{n_1+1}\epsilon^{n_1+2} + \epsilon^{n_1+1}\tau^{n_1+2} + \tau^{n_1}\epsilon^{n_1+3} + \epsilon^{n_1}\tau^{n_1+3}\right)$$
$$\geq 16\left(\tau^{n_1+1}\epsilon^{n_1+3} + \epsilon^{n_1+1}\tau^{n_1+3}\right) + \left(\tau^{n_1}\epsilon^{n_1+2} + \epsilon^{n_1}\tau^{n_1+2}\right).$$

If we take out the common factor of $\tau^{n_1}\epsilon^{n_1}$ then we get that this is equivalent to

$$\begin{aligned} 4\left(\tau\epsilon^2 + \epsilon\tau^2 + \epsilon^3 + \tau^3\right) &\geq 16\left(\tau\epsilon^3 + \epsilon\tau^3\right) + \epsilon^2 + \tau^2, \\ \iff 4(\epsilon^2 + \tau^2)(\epsilon + \tau) &\geq 16\tau\epsilon(\tau^2 + \epsilon^2) + (\epsilon^2 + \tau^2) \\ \iff 4(\epsilon + \tau) &\geq 16\tau\epsilon + 1, \end{aligned}$$

substituting $\epsilon = 1 - \tau$ gives that this is equivalent to

$$4 \geq 16\tau(1 - \tau) + 1$$
$$0 \geq -16\tau^2 + 16\tau - 3$$

i.e. the same identity as before. Therefore, by following the same reasoning as above, we can find counter-examples to NPIR in the range $\tau \in \left(\frac{1}{2}, \frac{3}{4}\right)$. Hence this last unresolved case has been resolved.

3.4.4 A Quick Note on SPIR & Ex

Within all of this you may well ask (since $Ex \rightarrow PIR$) where Constant Exchangeability fits in. It turns out that here we can find a counter-example to SPIR (and therefore, by corollary 3.2.4, to NPIR) for any given $\vec{n}$, as follows.

Take SPIR to be defined as

$$\frac{w(\langle n_1 + 2, n_2, n_3, \ldots, n_{2^p}\rangle)}{w(\langle n_1 + 1, n_2, n_3, \ldots, n_{2^p}\rangle)} \geq \frac{w(\langle n_1 + 1, n_2 + 1, n_3, \ldots, n_{2^p}\rangle)}{w(\langle n_1, n_2 + 1, n_3, \ldots, n_{2^p}\rangle)}, \tag{3.37}$$

and define a (non-symmetric) probability function w to be that with the de Finetti prior that assigns weight $\frac{1}{2}$ to each of the points[18]

$$\begin{aligned}\vec{a} &= \left\langle \frac{1}{2}, \tau, x, x, \ldots x \right\rangle \in \mathbb{D}_{2^p},\\ \vec{b} &= \left\langle \frac{1}{4}, \epsilon, y, y, \ldots y \right\rangle \in \mathbb{D}_{2^p},\end{aligned}$$

where $x = \frac{\frac{1}{2}-\tau}{2^p-2}$, $y = \frac{\frac{3}{4}-\epsilon}{2^p-2}$ and[19] $\tau \gg \epsilon$. Let $N = \sum_{i=3}^{2^p} n_i$, then we have that

$$\begin{aligned}w(\langle n_1+2, n_2, n_3 \ldots, n_{2^p}\rangle) &= \frac{1}{2}\left(\frac{1}{2}^{n_1+2}\tau^{n_2}x^N + \frac{1}{4}^{n_1+2}\epsilon^{n_2}y^N\right),\\ w(\langle n_1+1, n_2, n_3 \ldots, n_{2^p}\rangle) &= \frac{1}{2}\left(\frac{1}{2}^{n_1+1}\tau^{n_2}x^N + \frac{1}{4}^{n_1+1}\epsilon^{n_2}y^N\right),\\ w(\langle n_1+1, n_2+1, n_3 \ldots, n_{2^p}\rangle) &= \frac{1}{2}\left(\frac{1}{2}^{n_1+1}\tau^{n_2+1}x^N + \frac{1}{4}^{n_1+1}\epsilon^{n_2+1}y^N\right),\\ w(\langle n_1, n_2+1, n_3 \ldots, n_{2^p}\rangle) &= \frac{1}{2}\left(\frac{1}{2}^{n_1}\tau^{n_2+1}x^N + \frac{1}{4}^{n_1}\epsilon^{n_2+1}y^N\right).\end{aligned}$$

Plugging these into (3.37) gives

$$\frac{\frac{1}{2}^{n_1+2}\tau^{n_2}x^N + \frac{1}{4}^{n_1+2}\epsilon^{n_2}y^N}{\frac{1}{2}^{n_1+1}\tau^{n_2}x^N + \frac{1}{4}^{n_1+1}\epsilon^{n_2}y^N} \geq \frac{\frac{1}{2}^{n_1+1}\tau^{n_2+1}x^N + \frac{1}{4}^{n_1+1}\epsilon^{n_2+1}y^N}{\frac{1}{2}^{n_1}\tau^{n_2+1}x^N + \frac{1}{4}^{n_1}\epsilon^{n_2+1}y^N}$$

$$\Longleftrightarrow$$

$$\begin{aligned}&\frac{1}{2}^{2n_1+2}\tau^{2n_2+1}x^{2N} + \frac{1}{2}^{3n_1+4}\epsilon^{n_2}\tau^{n_2+1}x^Ny^N + \frac{1}{2}^{3n_1+2}\epsilon^{n_2+1}\tau^{n_2}x^Ny^N + \frac{1}{2}^{4n_1+4}\epsilon^{2n_2+1}y^{2N}\\ &\quad\geq \frac{1}{2}^{2n_1+2}\tau^{2n_2+1}x^{2N} + \frac{1}{2}^{3n_1+3}\epsilon^{n_2+1}\tau^{n_2}x^Ny^N + \frac{1}{2}^{3n_1+3}\epsilon^{n_2}\tau^{n_2+1}x^Ny^N\\ &\qquad + \frac{1}{2}^{4n_1+4}\epsilon^{2n_2+1}y^{2N}\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}&x^Ny^N\left(\frac{1}{2}^{3n_1+4}\epsilon^{n_2}\tau^{n_2+1} + \frac{1}{2}^{3n_1+2}\epsilon^{n_2+1}\tau^{n_2}\right)\\ &\quad\geq x^Ny^N\left(\frac{1}{2}^{3n_1+3}\epsilon^{n_2+1}\tau^{n_2} + \frac{1}{2}^{3n_1+3}\epsilon^{n_2}\tau^{n_2+1}\right)\end{aligned}$$

$$\begin{aligned}\Longleftrightarrow \frac{1}{2}^{3n_1+4}\epsilon^{n_2}\tau^{n_2}(\tau+4\epsilon) &\geq \frac{1}{2}^{3n_1+4}\epsilon^{n_2}\tau^{n_2}(2\epsilon+2\tau)\\ \Longleftrightarrow \tau+4\epsilon &\geq 2\epsilon+2\tau\\ \Longleftrightarrow 2\epsilon &\geq \tau,\end{aligned}$$

and so we can find counter-examples whenever $\tau > 2\epsilon$.

[18]Note that we are no longer dealing with SM-functions!
[19]We will discover how much larger τ has to be later on in this derivation.

3.4.5 Discussion

This is, of course, not what we had expected. SPIR is an ostensibly sensible principle, in that we'd consider it unreasonable to have greater belief in an event given that we'd just observed a different event occurring compared with our belief had we observed that event itself.

This leads us to consider a number of things, firstly that our initial assumptions may be incorrect. Our initial assumptions are basically that our probability functions ought to be symmetric, i.e. to satisfy Ax, and whilst there has been plenty of debate over the validity of such an assumption (see [32], which considers some of the problems and difficulties related to this area through historical examples) we regard it as unreasonable to relax.

Whilst Ax is considered our starting point, it is evident that we need to narrow down our range of candidate probability functions by using other natural principles - for instance JSP, which leads to Carnap's Continuum, or GPIR (leading to the Paris-Nix Continuum) - certainly the counter-examples we have found seem to relate to unusual and contrived situations so can SPIR be used in this way?

Unfortunately we were not able to find a characterisation which would allow us to state a continuum of probability functions defined by the fact that they satisfy SPIR. The utility of such a continuum, should it exist, is also questionable since it would appear to be very inclusive (for instance, it would have to include all of the two previously mentioned continua) though we can be hopeful that it would avoid triviality since it would exclude such apparently undesirable probability functions as those stated in our counter-examples.

Another problem with the possibility of using SPIR in this way is that there are other variants (such as NPIR) and it is difficult to justify exactly which variants we should use!

Worryingly (as far as the utility of SPIR is concerned) we have to ask why the strengthening of belief offered by SPIR holds in the special cases given in Theorem 3.4.2 but not in other cases which appear to be essentially the same? To this question we can, unfortunately, offer no convincing explanation.

3.5 PIR and Ex

We now switch our attention to the Principle of Instantial Relevance itself.

3.5.1 Historical Context

Rudolf Carnap first introduced the principle of PIR in chapter VI of [3], though he did so as an axiom, only reverting to calling it a principle in [6] after it had been proven to follow from Ex, more on which shortly. In these earlier works, Carnap also persisted in concentrating on the positive relevance aspect of the principle, i.e. when the $\geq$ inequality was replaced by $>$, but by [6] he had reverted to the principle as we know it, albeit with an additional rider that if the RHS was not equal to 0 or 1, then the principle held strictly.

It was Haim Gaifman, at the time a research student of Carnap, who first showed that PIR followed from Ex after following up on a suggestion[20] made to him by John Kemeny a year earlier, see [13]. Three years later, and independently of this previous work, Jürgen Humburg proved it again, this time using de Finetti's theorem in what Gaifman (in [13]) himself described as a "Cleverer Way", to produce a much shorter proof, see [19].

It does appear, however, that no-one has considered the converse, i.e. does PIR imply Ex? It is the intention of this section to rectify this.

[20]The suggestion was to try and use de Finetti's theorem.

3.5.2 PIR does not imply Ex

The version of PIR that we will initially be concerning ourselves with is, where $\theta \in SL$ is some state description, $\phi \in SL$ is some sentence and $\{a_{i1}, \ldots, a_{ih}\}$, $\{a_{j1}, \ldots, a_{jh}\}$, $\{a_{r1}, \ldots, a_{rg}\}$ are all disjoint,

$$w(\theta(a_{i1}, \ldots, a_{ih}) \mid \theta(a_{j1}, \ldots, a_{jh}) \wedge \phi(a_{r1}, \ldots, a_{rg})) \\ \geq w(\theta(a_{i1}, \ldots, a_{ih}) \mid \phi(a_{r1}, \ldots, a_{rg})). \tag{3.38}$$

This clearly implies the version of PIR defined at the beginning of the chapter, though (as we will later see) it is not the most general form of the principle.

In what follows we take L to be the language that contains a single predicate, P.

Lemma 3.5.1. *Take $\theta, \phi \in SL$, with no constants in common, and w to be the probability function that is defined by the de Finetti measure, μ, that puts weight 1 on the point $\langle x_1, x_2 \rangle \in \mathbb{D}_2$, then*

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

Proof. By the Disjunctive Normal Form Theorem (DNFT) we can assume that, for some r

$$\theta = \bigvee_{i=1}^{r} \langle n_{i1}, n_{i2} \rangle,$$

where $\langle n_{i1}, n_{i2} \rangle$ are vector representations of state descriptions. Similarly define

$$\phi = \bigvee_{j=1}^{t} \langle m_{j1}, m_{j2} \rangle,$$

for some t and vector representations $\langle m_{j1}, m_{j2} \rangle$.

Then

$$\begin{aligned}
w(\theta \wedge \phi) &= w\left(\bigvee_{i=1}^{r}\langle n_{i1}, n_{i2}\rangle \wedge \bigvee_{j=1}^{t}\langle m_{j1}, m_{j2}\rangle\right) \\
&= w\left(\bigvee_{i=1}^{r}\bigvee_{j=1}^{t}\langle n_{i1}+m_{j1}, n_{i2}+m_{j2}\rangle\right) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{t} w\left(\langle n_{i1}+m_{j1}, n_{i2}+m_{j2}\rangle\right) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{t} a_1^{n_{i1}+m_{j1}} a_2^{n_{i2}+m_{j2}} \\
&= \sum_{i=1}^{r}\sum_{j=1}^{t} a_1^{n_{i1}} a_1^{m_{j1}} a_2^{n_{i2}} a_2^{m_{j2}} \\
&= \left(\sum_{i=1}^{r} a_1^{n_{i1}} a_2^{n_{i2}}\right) \cdot \left(\sum_{j=1}^{t} a_1^{m_{j1}} a_2^{m_{j2}}\right) \\
&= \sum_{i=1}^{r} w\left(\langle n_{i1}, n_{i2}\rangle\right) \cdot \sum_{j=1}^{t} w\left(\langle m_{j1}, m_{j2}\rangle\right) \\
&= w\left(\bigvee_{i=1}^{r}\langle n_{i1}, n_{i2}\rangle\right) \cdot w\left(\bigvee_{j=1}^{t}\langle m_{j1}, m_{j2}\rangle\right) \\
&= w(\theta) \cdot w(\phi),
\end{aligned}$$

as required. □

Take $w = w^{\delta}$ where $\delta = \frac{1}{2}$. This means that $w = \frac{w^{+}+w^{-}}{2}$ where w^{+}, w^{-} are the probability functions which have de Finetti measures μ_{+}, μ_{-} that put weight 1 on the points $\langle \frac{3}{4}, \frac{1}{4}\rangle$ and $\langle \frac{1}{4}, \frac{3}{4}\rangle$, respectively. Set

$$v(P(a_1)) = v(\neg P(a_1)) = \frac{1}{2},$$

$$v(\theta(a_2, \ldots, a_n)) = w(\theta'(a_2, \ldots, a_{3(n-1)+1})),$$

where θ' is the result of replacing $P(a_j)$ everywhere by

$$\bigvee_{\epsilon_1+\epsilon_2+\epsilon_3 \geq 2} P^{\epsilon_1}(a_{3j-4}) \wedge P^{\epsilon_2}(a_{3j-3}) \wedge P^{\epsilon_3}(a_{3j-2}),$$

and replacing $\neg P(a_j)$ by

$$\bigvee_{\epsilon_1+\epsilon_2+\epsilon_3 < 2} P^{\epsilon_1}(a_{3j-4}) \wedge P^{\epsilon_2}(a_{3j-3}) \wedge P^{\epsilon_3}(a_{3j-2}),$$

where, $\epsilon_i \in \{0,1\}$ and P^0 represents $\neg P$, for each $j > 1$, note that any term involving the constant a_1 remains unchanged. Finally we set

$$v\left(P^{\epsilon}(a_1) \wedge \theta(a_2, a_3, \ldots, a_n)\right) = v(P^{\epsilon}(a_1)) \cdot v(\theta(a_2, a_3, \ldots, a_n)).$$

Our aim will be to show that v satisfies (3.38) but does not satisfy Ex, thus providing a counter-example to $PIR \rightarrow Ex$.

Given a state description

$$\theta = P^{\epsilon}(a_1) \wedge \langle n, m \rangle,$$

let

$$\psi = \langle n, m \rangle,$$

then we have that

$$\psi' = \bigwedge_n \left(\langle 3,0 \rangle \vee \bigvee_3 \langle 2,1 \rangle \right) \wedge \bigwedge_m \left(\langle 0,3 \rangle \vee \bigvee_3 \langle 1,2 \rangle \right).$$

Next note that

$$\begin{aligned}
w^+\left(\langle 3,0 \rangle \vee \bigvee_3 \langle 2,1 \rangle \right) &= \left(\left(\frac{3}{4}\right)^3 + 3 \cdot \frac{1}{4} \left(\frac{3}{4}\right)^2 \right) = \frac{27}{32}, \\
w^+\left(\langle 0,3 \rangle \vee \bigvee_3 \langle 1,2 \rangle \right) &= \left(\left(\frac{1}{4}\right)^3 + 3 \cdot \left(\frac{1}{4}\right)^2 \frac{3}{4} \right) = \frac{5}{32}, \\
w^-\left(\langle 3,0 \rangle \vee \bigvee_3 \langle 2,1 \rangle \right) &= w^+\left(\langle 0,3 \rangle \vee \bigvee_3 \langle 1,2 \rangle \right) = \frac{5}{32}, \\
w^-\left(\langle 0,3 \rangle \vee \bigvee_3 \langle 1,2 \rangle \right) &= w^+\left(\langle 3,0 \rangle \vee \bigvee_3 \langle 2,1 \rangle \right) = \frac{27}{32}.
\end{aligned}$$

By Lemma 3.5.1 we have that,

$$\begin{aligned}
w^+\left(\bigwedge_n \left(\langle 3,0 \rangle \vee \bigvee_3 \langle 2,1 \rangle \right) \right) &= \left(\left(\frac{3}{4}\right)^3 + 3 \cdot \frac{1}{4} \left(\frac{3}{4}\right)^2 \right)^n = \left(\frac{27}{32}\right)^n, \\
w^+\left(\bigwedge_m \left(\langle 0,3 \rangle \vee \bigvee_3 \langle 1,2 \rangle \right) \right) &= \left(\frac{5}{32}\right)^m, \\
w^-\left(\bigwedge_n \left(\langle 3,0 \rangle \vee \bigvee_3 \langle 2,1 \rangle \right) \right) &= \left(\frac{5}{32}\right)^n, \\
w^-\left(\bigwedge_m \left(\langle 0,3 \rangle \vee \bigvee_3 \langle 1,2 \rangle \right) \right) &= \left(\frac{27}{32}\right)^m.
\end{aligned}$$

Again by Lemma 3.5.1 we have that,

$$\begin{aligned} w^+(\psi') &= w^+\left(\bigwedge_n \left(\langle 3,0\rangle \vee \bigvee_3 \langle 2,1\rangle\right) \wedge \bigwedge_m \left(\langle 0,3\rangle \vee \bigvee_3 \langle 1,2\rangle\right)\right) \\ &= \left(\frac{27}{32}\right)^n \left(\frac{5}{32}\right)^m, \\ w^-(\psi') &= \left(\frac{5}{32}\right)^n \left(\frac{27}{32}\right)^m. \end{aligned}$$

So overall we have that:

$$v(\psi) = w(\psi') = \frac{1}{2}\left(\left(\frac{27}{32}\right)^n \left(\frac{5}{32}\right)^m + \left(\frac{5}{32}\right)^n \left(\frac{27}{32}\right)^m\right),$$

and,

$$v(\theta) = v(P^\epsilon(a_1)) \cdot w(\psi') = \frac{1}{4}\left(\left(\frac{27}{32}\right)^n \left(\frac{5}{32}\right)^m + \left(\frac{5}{32}\right)^n \left(\frac{27}{32}\right)^m\right).$$

These facts will make the calculations much easier in what follows. We are now in a position to state and prove the main result of the section.

Theorem 3.5.2. *The probability function v, as defined above, satisfies (3.38) but fails to satisfy Ex.*

Proof. We start by showing that v fails Ex. Consider

$$\begin{aligned} v\left(\neg P(a_1) \wedge P(a_2) \wedge P(a_3)\right) &= v\left(\neg P(a_1) \wedge \langle 2,0\rangle\right) \\ &= \frac{1}{4}\left(\left(\frac{27}{32}\right)^2 + \left(\frac{5}{32}\right)^2\right) \\ &= \frac{1}{4}\left(\frac{729}{1024} + \frac{25}{1024}\right) \\ &= \frac{377}{2048}. \end{aligned}$$

If v satisfies Ex then by permuting a_1 and a_2 we should get that v awards an equal probability (of $\frac{377}{2048}$) to this new sentence. However

$$\begin{aligned} v\left(\neg P(a_2) \wedge P(a_1) \wedge P(a_3)\right) &= v\left(P(a_1) \wedge \langle 1,1\rangle\right) \\ &= \frac{1}{4}\left(\frac{27}{32}\frac{5}{32} + \frac{27}{32}\frac{5}{32}\right) \\ &= \frac{1}{2}\left(\frac{27}{32} \cdot \frac{5}{32}\right) \\ &= \frac{135}{2048}, \end{aligned}$$

thus a counter-example is formed and v fails to satisfy Ex.

We now show that v satisfies (3.38). We consider the following expanded version of (3.38);

$$\begin{aligned} & v(\theta(\vec{a_i}) \wedge \theta(\vec{a_j}) \wedge \phi(\vec{a_r})) \cdot v(\phi(\vec{a_r})) \\ & \qquad \geq v(\theta(\vec{a_i}) \wedge \phi(\vec{a_r})) \cdot v(\theta(\vec{a_j}) \wedge \phi(\vec{a_r})), \end{aligned} \tag{3.39}$$

where $\vec{a_i} = \langle a_{i1}, \ldots, a_{ih} \rangle$, $\vec{a_j} = \langle a_{j1}, \ldots, a_{jh} \rangle$, $\vec{a_r} = \langle a_{r1}, \ldots, a_{rg} \rangle$, $\theta, \phi \in SL$ with θ a state description. We must consider several cases, based upon whether a_1 appears in $\vec{a_i}$, $\vec{a_j}$ or neither.

Case 1 : a_1 does not appear in $\vec{a_i}$ or $\vec{a_j}$.
Note that if a_1 also does not appear in $\vec{a_r}$ then the result is trivial, since v satisfies Ex if it is only handling sentences defined over a subset of constants not including a_1.

Let θ be the state description with the vector representation of $\langle n, m \rangle$ and take $\vec{a_i'}$ to represent the range of constants that θ' is defined over when transformed from $\theta(\vec{a_i})$, similarly define $\vec{a_j'}$ and $\vec{a_r'}$, then we have that

$$\begin{aligned} w^+(\theta'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})) &= \left(\frac{27}{32}\right)^{2n} \left(\frac{5}{32}\right)^{2m} w^+(\phi'), \\ w^+(\theta'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})) &= \left(\frac{27}{32}\right)^{n} \left(\frac{5}{32}\right)^{m} w^+(\phi'(\vec{a_r'})) = w^+(\theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})), \\ w^-(\theta'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})) &= \left(\frac{27}{32}\right)^{2m} \left(\frac{5}{32}\right)^{2n} w^-(\phi'(\vec{a_r'})), \\ w^-(\theta'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})) &= \left(\frac{27}{32}\right)^{m} \left(\frac{5}{32}\right)^{n} w^-(\phi') = w^-(\theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})). \end{aligned}$$

If we take $n \geq m$ and apply these to (3.39) we get (noting that both $w^+(\phi')$ and $w^-(\phi')$ are non-negative):

$$\begin{aligned} & \frac{1}{2}\left(\left(\frac{27}{32}\right)^{2n} \left(\frac{5}{32}\right)^{2m} w^+(\phi') + \left(\frac{27}{32}\right)^{2m} \left(\frac{5}{32}\right)^{2n} w^-(\phi')\right) \cdot \frac{1}{2}\left(w^+(\phi') + w^-(\phi')\right) \\ & \quad \geq \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n} \left(\frac{5}{32}\right)^{m} w^+(\phi') + \left(\frac{27}{32}\right)^{m} \left(\frac{5}{32}\right)^{n} w^-(\phi')\right) \\ & \qquad \times \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n} \left(\frac{5}{32}\right)^{m} w^+(\phi') + \left(\frac{27}{32}\right)^{m} \left(\frac{5}{32}\right)^{n} w^-(\phi')\right) \end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m} w^+(\phi')^2 + \left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m} w^+(\phi')w^-(\phi') \\
&\quad + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n} w^+(\phi')w^-(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n} w^-(\phi')^2 \\
&\quad \geq \left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m} w^+(\phi')^2 + 2\cdot\left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m} w^+(\phi')w^-(\phi') \\
&\quad + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n} w^-(\phi')^2
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m} w^+(\phi')w^-(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n} w^+(\phi')w^-(\phi') \\
&\quad \geq 2\cdot\left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m} w^+(\phi')w^-(\phi')
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m} + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n} \\
&\quad \geq 2\cdot\left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m}
\end{aligned}$$

$$\Longleftrightarrow$$

$$\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m} + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n} - 2\cdot\left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m} \geq 0$$

$$\Longleftrightarrow$$

$$\left(\left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2m}\right)\left(\left(\frac{27}{32}\right)^{2(n-m)} + \left(\frac{5}{32}\right)^{2(n-m)} - 2\cdot\left(\frac{27}{32}\right)^{n-m}\left(\frac{5}{32}\right)^{n-m}\right) \geq 0 \tag{3.40}$$

$$\Longleftrightarrow$$

$$\left(\left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2m}\right)\left(\left(\frac{27}{32}\right)^{n-m} - \left(\frac{5}{32}\right)^{n-m}\right)^2 \geq 0,$$

and this last line holds trivially since both multiplicands on the LHS are non-negative. Note that if $m > n$ then we proceed as above until we reach (3.40) where we instead take out a common factor of

$$\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2n}\right),$$

to give that (3.39) is in this case is equivalent to

$$\begin{aligned}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2n}\right)\left(\left(\frac{27}{32}\right)^{2(m-n)}+\left(\frac{5}{32}\right)^{2(m-n)}-2\cdot\left(\frac{27}{32}\right)^{m-n}\left(\frac{5}{32}\right)^{m-n}\right) &\geq 0\\ \iff \left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2n}\right)\left(\left(\frac{27}{32}\right)^{m-n}-\left(\frac{5}{32}\right)^{m-n}\right)^{2} &\geq 0,\end{aligned}$$

and, again, we have that the last line holds since both multiplicands on the LHS are non-negative. Therefore the inequality holds, and PIR is satisfied for this case. As an aside, note that all the terms involving ϕ' cancel out, meaning that this result holds irrespective of the form of ϕ. This will also be the case for the sequel.

Case 2 : a_1 appears in $\vec{a_i}$.

There are two subcases to consider, these are:

(a) When $\theta(\vec{a_i})$ contains $P(a_1)$,

(b) When $\theta(\vec{a_i})$ contains $\neg P(a_1)$.

Throughout the following let θ be a state description with vector representation $\langle n, m\rangle$. We start with subcase (a). Here we have that

$$\theta(\vec{a_i}) = P(a_1) \wedge \langle n-1, m\rangle.$$

Let

$$\psi(\vec{a_i} - a_1) = \langle n-1, m\rangle,$$

so,

$$\begin{aligned}
w^{+}\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi'),\\
w^{+}\left(\psi'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^{+}(\phi'),\\
w^{+}\left(\theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi'),
\end{aligned}$$

$$\begin{aligned}
w^{-}\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi'),\\
w^{-}\left(\psi'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^{-}(\phi'),\\
w^{-}\left(\theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi').
\end{aligned}$$

Therefore

$$\begin{aligned}
&v\left(P(a_1) \wedge \psi(\vec{a_i} - a_1) \wedge \theta(\vec{a_j}) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi')\right),\\
&v\left(P(a_1) \wedge \psi(\vec{a_i} - a_1) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^{-}(\phi')\right),\\
&v\left(\theta(\vec{a_j}) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi')\right).
\end{aligned}$$

If we take $n-1 \geq m$ and apply these to (3.39) we get

$$\begin{aligned}
&\frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi')\right)\\
&\qquad \times \frac{1}{2}\left(w^{+}(\phi') + w^{-}(\phi')\right)\\
&\quad \geq \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^{-}(\phi')\right)\\
&\qquad\quad \times \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi')\right)
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\frac{1}{8}\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi')\right)\left(w^{+}(\phi') + w^{-}(\phi')\right) \\
&\quad\geq\ \frac{1}{8}\left(\left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^{-}(\phi')\right) \\
&\qquad\qquad \times\ \left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi')\right)
\end{aligned} \tag{3.41}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi')^2 \ +\ \left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi')w^{-}(\phi') \\
&\quad +\ \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{+}(\phi')w^{-}(\phi')\ +\ \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi')^2 \\
&\geq\ \left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi')^2 \ +\ \left(\frac{27}{32}\right)^{n+m-1}\left(\frac{5}{32}\right)^{n+m} w^{+}(\phi')w^{-}(\phi') \\
&\quad +\ \left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m-1} w^{+}(\phi')w^{-}(\phi') \\
&\quad +\ \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi')^2
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi')w^{-}(\phi')\ +\ \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{+}(\phi')w^{-}(\phi') \\
&\quad\geq\ \left(\frac{27}{32}\right)^{n+m-1}\left(\frac{5}{32}\right)^{n+m} w^{+}(\phi')w^{-}(\phi') \\
&\qquad +\ \left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m-1} w^{+}(\phi')w^{-}(\phi')
\end{aligned} \tag{3.42}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} \\
&\quad\geq\ \left(\frac{27}{32}\right)^{n+m-1}\left(\frac{5}{32}\right)^{n+m} + \left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m-1}
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} \\
&\quad -\left(\frac{27}{32}\right)^{n+m-1}\left(\frac{5}{32}\right)^{n+m} - \left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m-1}\ \geq\ 0
\end{aligned}$$

$$\Longleftrightarrow$$

$$\left(\left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2m}\right)\left(\left(\frac{27}{32}\right)^{2n-1-2m}+\left(\frac{5}{32}\right)^{2n-1-2m}\right.$$
$$\left.-\left(\frac{27}{32}\right)^{n-1-m}\left(\frac{5}{32}\right)^{n-m}-\left(\frac{27}{32}\right)^{n-m}\left(\frac{5}{32}\right)^{n-1-m}\right)\geq 0 \tag{3.43}$$

$$\Longleftrightarrow$$

$$\left(\left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2m}\right)$$
$$\times\left(\left(\frac{27}{32}\right)^{n-1-m}-\left(\frac{5}{32}\right)^{n-1-m}\right)\left(\left(\frac{27}{32}\right)^{n-m}-\left(\frac{5}{32}\right)^{n-m}\right)\geq 0. \tag{3.44}$$

All three multiplicands in (3.44) are positive and therefore the inequality holds, as required. Note that if we had taken $m > n-1$ then the derivation would have followed exactly as above until (3.43) where we instead take out a common factor of

$$\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2n-1}\right),$$

to give that (3.39) is in this case is equivalent to

$$\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2n-1}\right)\left(\left(\frac{27}{32}\right)^{2m-2n+1}+\left(\frac{5}{32}\right)^{2m-2n+1}\right.$$
$$\left.-\left(\frac{27}{32}\right)^{m-n}\left(\frac{5}{32}\right)^{m-n+1}-\left(\frac{27}{32}\right)^{m-n+1}\left(\frac{5}{32}\right)^{m-n}\right)\geq 0$$

$$\Longleftrightarrow$$

$$\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2n-1}\right)$$
$$\times\left(\left(\frac{27}{32}\right)^{m-n+1}-\left(\frac{5}{32}\right)^{m-n+1}\right)\left(\left(\frac{27}{32}\right)^{m-n}-\left(\frac{5}{32}\right)^{m-n}\right)\geq 0,$$

and again all three multiplicands are positive and so the inequality holds, as required. Case 2a is therefore shown.

For case 2b we have that

$$\theta(\vec{a_i}) = \neg P(a_1) \wedge \langle n, m-1 \rangle.$$

This time we take

$$\psi(\vec{a_i} - a_1) = \langle n, m-1 \rangle,$$

and so;

$$\begin{aligned}
w^+\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi'),\\
w^+\left(\psi'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m-1} w^+(\phi'),\\
w^+\left(\theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi'),
\end{aligned}$$

$$\begin{aligned}
w^-\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^-(\phi'),\\
w^-\left(\psi'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n} w^-(\phi'),\\
w^-\left(\theta'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \frac{27}{32}^{m}\frac{5}{32}^{n} w^-(\phi').
\end{aligned}$$

Therefore

$$\begin{aligned}
&v\left(\neg P(a_1) \wedge \psi(\vec{a_i} - a_1) \wedge \theta(\vec{a_j}) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi') + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^-(\phi')\right),\\
&v\left(\neg P(a_1) \wedge \psi(\vec{a_i} - a_1) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m-1} w^+(\phi') + \left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right),\\
&v\left(\theta(\vec{a_j}) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \frac{27}{32}^{m}\frac{5}{32}^{n} w^-(\phi')\right).
\end{aligned}$$

Applying these to (3.39) gives

$$\begin{aligned}
&\frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi') + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^-(\phi')\right) \\
&\qquad \times \frac{1}{2}\left(w^+(\phi') + w^-(\phi')\right) \\
&\qquad \geq \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m-1} w^+(\phi') + \left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right) \\
&\qquad\qquad \times \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \frac{27}{32}^{m}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right)
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\frac{1}{8}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi') + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^-(\phi')\right)\left(w^+(\phi') + w^-(\phi')\right) \\
&\quad \geq \frac{1}{8}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m-1} w^+(\phi') + \left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right) \\
&\qquad \times \left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \frac{27}{32}^{m}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right)
\end{aligned} \tag{3.45}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi')^2 + \left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi')w^-(\phi') \\
&\quad + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^+(\phi')w^-(\phi') + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^-(\phi')^2 \\
&\quad \geq \left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi')^2 + \left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m-1} w^+(\phi')w^-(\phi') \\
&\qquad + \left(\frac{27}{32}\right)^{n+m-1}\left(\frac{5}{32}\right)^{n+m} w^+(\phi')w^-(\phi') \\
&\qquad\quad + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^-(\phi')^2
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
&\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^+(\phi')w^-(\phi') + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^+(\phi')w^-(\phi') \\
&\quad \geq \left(\frac{27}{32}\right)^{n+m}\left(\frac{5}{32}\right)^{n+m-1} w^+(\phi')w^-(\phi') \\
&\qquad + \left(\frac{27}{32}\right)^{n+m-1}\left(\frac{5}{32}\right)^{n+m} w^+(\phi')w^-(\phi').
\end{aligned}$$

This last inequality is equivalent to (3.42) except with the n's and m's switched. The equivalent derivation as given there gives that (3.46) is equivalent to

$$\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2n}\right) \times \left(\left(\frac{27}{32}\right)^{m-1-n} - \left(\frac{5}{32}\right)^{m-1-n}\right)\left(\left(\frac{27}{32}\right)^{m-n} - \left(\frac{5}{32}\right)^{m-n}\right) \geq 0, \tag{3.46}$$

when $m-1 \geq n$ and

$$\left(\left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2m-1}\right) \times \left(\left(\frac{27}{32}\right)^{n-m+1} - \left(\frac{5}{32}\right)^{n-m+1}\right)\left(\left(\frac{27}{32}\right)^{n-m} - \left(\frac{5}{32}\right)^{n-m}\right) \geq 0, \tag{3.47}$$

otherwise. Both (3.46) and (3.47) have that the three multiplicands on the LHS are all positive, and therefore both inequalities hold and Case 2b is shown, as required.

Case 3 : a_1 appears in $\vec{a_j}$.

As with case 2 we have to consider two subcases, these are:

(a) When $\theta(\vec{a_j})$ contains $P(a_1)$,

(b) When $\theta(\vec{a_j})$ contains $\neg P(a_1)$.

We start with subcase (a). Here we have that

$$\theta(\vec{a_j}) = P(a_1) \wedge \langle n-1, m\rangle.$$

Let

$$\psi(\vec{a_j} - a_1) = \langle n-1, m\rangle$$

so;

$$\begin{aligned}
w^+\left(\theta'(\vec{a_i'}) \wedge \psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^+(\phi'),\\
w^+\left(\theta'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi'),\\
w^+\left(\psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^+(\phi'),
\end{aligned}$$

$$\begin{aligned}
w^-\left(\theta'(\vec{a_i'}) \wedge \psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^-(\phi'),\\
w^-\left(\theta'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^-(\phi'),\\
w^-\left(\psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^-(\phi').
\end{aligned}$$

Therefore

$$\begin{aligned}
&v\left(\theta(\vec{a_i}) \wedge P(a_1) \wedge \psi(\vec{a_j} - a_1) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^+(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^-(\phi')\right),\\
&v\left(\theta(\vec{a_i}) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right),\\
&v\left(P(a_1) \wedge \psi(\vec{a_j} - a_1) \wedge \phi(\vec{a_r})\right)\\
&\qquad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \frac{27}{32}^{m}\frac{5}{32}^{n-1} w^-(\phi')\right).
\end{aligned}$$

When we plug these into (3.39) we get

$$\begin{aligned}
&\frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^+(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^-(\phi')\right)\\
&\qquad \times \frac{1}{2}\left(w^+(\phi') + w^-(\phi')\right)\\
&\quad \geq \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^-(\phi')\right)\\
&\qquad\qquad \times \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^+(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^-(\phi')\right),
\end{aligned}$$

$$\Longleftrightarrow$$

$$
\begin{aligned}
&\frac{1}{8}\left(\left(\frac{27}{32}\right)^{2n-1}\left(\frac{5}{32}\right)^{2m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{2m}\left(\frac{5}{32}\right)^{2n-1} w^{-}(\phi')\right)\left(w^{+}(\phi') + w^{-}(\phi')\right) \\
&\quad \geq \frac{1}{8}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi')\right) \\
&\qquad\quad \times \left(\left(\frac{27}{32}\right)^{n-1}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n-1} w^{-}(\phi')\right),
\end{aligned}
$$

which is equivalent to (3.41). The proof of this case then follows exactly as for case 2a.

We now consider case 3b. Here we have that

$$\theta(\vec{a_j}) = \neg P(a_1) \wedge \langle n, m-1 \rangle.$$

Let

$$\psi(\vec{a_j} - a_1) = \langle n, m-1 \rangle,$$

so;

$$
\begin{aligned}
w^{+}\left(\theta'(\vec{a_i'}) \wedge \psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^{+}(\phi'), \\
w^{+}\left(\theta'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi'), \\
w^{+}\left(\psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m-1} w^{+}(\phi'), \\
w^{-}\left(\theta'(\vec{a_i'}) \wedge \psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^{-}(\phi'), \\
w^{-}\left(\theta'(\vec{a_i'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi'), \\
w^{-}\left(\psi'(\vec{a_j'}) \wedge \phi'(\vec{a_r'})\right) &= \left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n} w^{-}(\phi').
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&v\left(\theta(\vec{a_i}) \wedge \neg P(a_1) \wedge \psi(\vec{a_j} - a_1) \wedge \phi(\vec{a_r})\right) \\
&\quad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1} w^{+}(\phi') + \left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n} w^{-}(\phi')\right), \\
&v\left(\theta(\vec{a_i}) \wedge \phi(\vec{a_r})\right) \\
&\quad = \frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m} w^{+}(\phi') + \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n} w^{-}(\phi')\right), \\
&v\left(\neg P(a_1) \wedge \psi(\vec{a_j} - a_1) \wedge \phi(\vec{a_r})\right) \\
&\quad = \frac{1}{4}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m-1} w^{+}(\phi') + \frac{27}{32}^{m-1}\frac{5}{32}^{n} w^{-}(\phi')\right).
\end{aligned}
$$

Applying these to (3.39) gives

$$\frac{1}{4}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1}w^{+}(\phi')+\left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n}w^{-}(\phi')\right)$$
$$\times\frac{1}{2}\left(w^{+}(\phi')+w^{-}(\phi')\right)$$
$$\geq\frac{1}{2}\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m}w^{+}(\phi')+\left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n}w^{-}(\phi')\right)$$
$$\times\frac{1}{4}\left(\left(\frac{27^{n}}{32}\right)\left(\frac{5}{32}\right)^{m-1}w^{+}(\phi')+\left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n}w^{-}(\phi')\right),$$

$$\Longleftrightarrow$$

$$\frac{1}{8}\left(\left(\frac{27}{32}\right)^{2n}\left(\frac{5}{32}\right)^{2m-1}w^{+}(\phi')+\left(\frac{27}{32}\right)^{2m-1}\left(\frac{5}{32}\right)^{2n}w^{-}(\phi')\right)\left(w^{+}(\phi')+w^{-}(\phi')\right)$$
$$\geq\left(\left(\frac{27}{32}\right)^{n}\left(\frac{5}{32}\right)^{m}w^{+}(\phi')+\left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{n}w^{-}(\phi')\right)$$
$$\times\left(\left(\frac{27^{n}}{32}\right)\left(\frac{5}{32}\right)^{m-1}w^{+}(\phi')+\left(\frac{27}{32}\right)^{m-1}\left(\frac{5}{32}\right)^{n}w^{-}(\phi')\right),$$

which is equivalent to (3.45). The proof of this case then follows exactly as for case 2b.

All cases are therefore proven, as required. $\square$

3.5.3 The Most General Form of PIR

So we have that $PIR \nrightarrow Ex$ with PIR as defined in (3.38), what happens if we instead take the most general form of PIR, given by

$$w\left(\theta(\vec{a_i}) \mid \theta(\vec{a_j}) \wedge \phi(\vec{a_r})\right) \geq w\left(\theta(\vec{a_i}) \mid \phi(\vec{a_r})\right), \tag{3.48}$$

where $\{a_{i1},\ldots,a_{ih}\}$, $\{a_{j1},\ldots,a_{jh}\}$, $\{a_{r1},\ldots,a_{rg}\}$ are all disjoint, and $\theta,\phi \in SL$ (i.e. this is the same definition as before except that we are now allowing θ to be any sentence as opposed to any state description) ?

We do not intend to answer this question fully, though we will show that the result of the last section does not carry over to this by showing that v does not satisfy (3.48).

We take

$$\theta(\vec{a_i}) = (P(a_1) \wedge \langle m+1, m\rangle) \vee \left(\neg P(a_1) \wedge \bigvee_2 \langle m, m\rangle\right),$$

where $m \geq 1$ and the two state descriptions involved in the disjunction

$$\bigvee_2 \langle m, m\rangle,$$

are defined over the same $2m$ constants and are distinct[21] . Take $\theta(\vec{a_j})$ to be the same only over a disjoint set of constants (i.e. $a_1 \notin \vec{a_j}$).

Set $\phi = P(a_s)$ where a_s is some constant not in $\vec{a_i}$ or $\vec{a_j}$.

Note that, because of the way in which we have defined θ, we can write

$$v(\theta(\vec{a_i})) = \frac{1}{2} \cdot v(\langle m+1, m\rangle \vee \bigvee_2 \langle m, m\rangle).$$

Let

$$\psi(\vec{a_i} - a_1) = \langle m+1, m\rangle \vee \bigvee_2 \langle m, m\rangle,$$

then we have that (3.48) is for v equivalent to (since w trivially satisfies REG);

$$\frac{\frac{1}{2} \cdot w\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(a_s')\right)}{w\left(\theta'(\vec{a_j'}) \wedge \phi'(a_s')\right)} \geq \frac{\frac{1}{2} \cdot w\left(\psi'(\vec{a_i'}) \wedge \phi'(a_s')\right)}{w\left(\phi'(a_s')\right)}$$

$$\iff$$

$$\frac{\frac{1}{4}\left(w^+\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(a's)\right) + w^-\left(\psi'(\vec{a_i'}) \wedge \theta'(\vec{a_j'}) \wedge \phi'(a's)\right)\right)}{\frac{1}{2}\left(w^+\left(\theta'(\vec{a_j'}) \wedge \phi'(a_s')\right) + w^-\left(\theta'(\vec{a_j'}) \wedge \phi'(a_s')\right)\right)}$$

$$\geq \frac{\frac{1}{4}\left(w^+\left(\psi'(\vec{a_i'}) \wedge \phi'(a_s')\right) + w^-\left(\psi'(\vec{a_i'}) \wedge \phi'(a_s')\right)\right)}{\frac{1}{2}\left(w^+\left(\phi'(a_s')\right) + w^-\left(\phi'(a_s')\right)\right)}. \quad (3.49)$$

[21]It is always possible to find two such state descriptions, for instance, by swapping the first observation of P with the first observation of $\neg P$. There must be at least one observation of each since $m \geq 1$.

We cancel the factor of $\frac{1}{2}$ and use Lemma 3.5.1 to give that (3.49) is equivalent to

$$\frac{w^+(\psi'(\vec{a'_i}))w^+(\theta'(\vec{a'_j}))w^+(\phi'(a'_s)) + w^-(\psi'(\vec{a'_i}))w^-(\theta'(\vec{a'_j}))w^-(\phi'(a'_s))}{w^+(\theta'(\vec{a'_j}))w^+(\phi'(a'_s)) + w^-(\theta'(\vec{a'_j}))w^-(\phi'(a'_s))}$$

$$\geq \frac{w^+(\psi'(\vec{a'_i}))w^+(\phi'(a'_s)) + w^-(\psi'(\vec{a'_i}))w^-(\phi'(a'_s))}{w^+(\phi'(a'_s)) + w^-(\phi'(a'_s))}. \tag{3.50}$$

Now, trivially

$$w^+(\phi'(a'_s)) + w^-(\phi'(a'_s)) = 1,$$

and so (3.50) is equivalent to

$$\frac{w^+(\psi'(\vec{a'_i}))w^+(\theta'(\vec{a'_j}))w^+(\phi'(a'_s)) + w^-(\psi'(\vec{a'_i}))w^-(\theta'(\vec{a'_j}))w^-(\phi'(a'_s))}{w^+(\theta'(\vec{a'_j}))w^+(\phi'(a'_s)) + w^-(\theta'(\vec{a'_j}))w^-(\phi'(a'_s))}$$

$$\geq w^+(\psi'(\vec{a'_i}))w^+(\phi'(a'_s)) + w^-(\psi'(\vec{a'_i}))w^-(\phi'(a'_s))$$

$$\iff$$

$$\begin{aligned} w^+(\psi'(\vec{a'_i}))w^+(\theta'(\vec{a'_j}))w^+(\phi'(a'_s)) &+ w^-(\psi'(\vec{a'_i}))w^-(\theta'(\vec{a'_j}))w^-(\phi'(a'_s)) \\ &\geq \Big(w^+(\theta'(\vec{a'_j}))w^+(\phi'(a'_s)) + w^-(\theta'(\vec{a'_j}))w^-(\phi'(a'_s))\Big) \\ &\quad \times \Big(w^+(\psi'(\vec{a'_i}))w^+(\phi'(a'_s)) + w^-(\psi'(\vec{a'_i}))w^-(\phi'(a'_s))\Big) \end{aligned} \tag{3.51}$$

and this is the condition that we will be testing.

$$\begin{aligned}
w^+(\psi'(\vec{a_i'})) &= \left(\left(\frac{27}{32}\right)^{m+1}\left(\frac{5}{32}\right)^m + 2\cdot\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right) \\
&= \left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)\left(\frac{27}{32}+2\right),
\end{aligned}$$

$$\begin{aligned}
w^-(\psi'(\vec{a_i'})) &= \left(\left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{m+1} + 2\cdot\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right) \\
&= \left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)\left(\frac{5}{32}+2\right),
\end{aligned}$$

$$\begin{aligned}
w^+(\theta'(\vec{a_j'})) &= \left(\frac{27}{32}\right)^{m+2}\left(\frac{5}{32}\right)^m + 2\cdot\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^{m+1} \\
&= \left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)\left(\left(\frac{27}{32}\right)^2+\frac{10}{32}\right),
\end{aligned}$$

$$\begin{aligned}
w^-(\theta'(\vec{a_j'})) &= \left(\frac{27}{32}\right)^{m}\left(\frac{5}{32}\right)^{m+2} + 2\cdot\left(\frac{27}{32}\right)^{m+1}\left(\frac{5}{32}\right)^{m} \\
&= \left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)\left(\left(\frac{5}{32}\right)^2+\frac{54}{32}\right),
\end{aligned}$$

$$w^+(\phi'(a_s')) = \frac{27}{32}, \quad w^-(\phi'(a_s')) = \frac{5}{32}.$$

plugging these values into (3.51) gives that

$$\begin{aligned}
&\left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)^2 \\
&\quad\times\left(\frac{27}{32}\left(\frac{27}{32}+2\right)\left(\left(\frac{27}{32}\right)^2+\frac{10}{32}\right)+\frac{5}{32}\left(\frac{5}{32}+2\right)\left(\left(\frac{5}{32}\right)^2+\frac{54}{32}\right)\right) \\
&\qquad\geq \left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)^2\left(\frac{27}{32}\left(\left(\frac{27}{32}\right)^2+\frac{10}{32}\right)+\frac{5}{32}\left(\left(\frac{5}{32}\right)^2+\frac{54}{32}\right)\right) \\
&\qquad\quad\times\left(\frac{27}{32}\left(\frac{27}{32}+2\right)+\frac{5}{32}\left(\frac{5}{32}+2\right)\right). \qquad (3.52)
\end{aligned}$$

Mathematica gives that (to 5 decimal places[22]):

$$\begin{aligned}
\frac{27}{32}\left(\frac{27}{32}+2\right)\left(\left(\frac{27}{32}\right)^2+\frac{10}{32}\right) &= 2.45799,\\
\frac{5}{32}\left(\frac{5}{32}+2\right)\left(\left(\frac{5}{32}\right)^2+\frac{54}{32}\right) &= 0.57677,\\
\left(\frac{27}{32}\left(\left(\frac{27}{32}\right)^2+\frac{10}{32}\right)+\frac{5}{32}\left(\left(\frac{5}{32}\right)^2+\frac{54}{32}\right)\right) &= 1.13184,\\
\left(\frac{27}{32}\left(\frac{27}{32}+2\right)+\frac{5}{32}\left(\frac{5}{32}+2\right)\right) &= 2.73633.
\end{aligned}$$

Plugging these values into (3.52) and removing the common factor of

$$\left(\left(\frac{27}{32}\right)^m\left(\frac{5}{32}\right)^m\right)^2,$$

gives that this is equivalent to

$$\begin{aligned}
2.45799+0.57677 &\geq 1.13184\times 2.73633\\
\Longleftrightarrow 3.03476 &\geq 3.09707,
\end{aligned}$$

and thus the required contradiction is formed, so the probability function v does not satisfy this version of PIR.

It seems as though our ability to find a counter-example to the notion of $PIR \to EX$ depends entirely upon how we define PIR. We believe that this result is not as big a problem as it may appear since the way in which we state the principle in (3.38) most closely captures the intuition behind the principle, i.e. that once we have observed something we then should not award a lower probability to observing it again than we did before. It is therefore suggested that the case that is of greater interest to us is exactly the case in which we can show that $PIR \nrightarrow EX$, though perhaps it would be of interest to clear up the latter case in the future.

[22]All calculations were carried out using exact figures, the results are given to 5dp for reasons of brevity.

Chapter 4

Principles Involving Irrelevance

4.1 Introduction & Motivation

We now switch our attention from principles based upon relevance to principles based upon *irrelevance*, traditionally a more difficult notion to pin down.

For the most part, this chapter is concerned with studying the Weak Irrelevance Principle (WIP), the weakest in a family including the Predicate (PIP), Constant (IP) and Conditional (CIP) irrelevance principles (introduced and discussed in [18]). WIP is defined as follows;

Definition 4.1.1. The probability function, w, satisfies WIP if for any two sentences, θ, $\phi \in SL$, that have no constants or predicates in common we have that

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

Note that this is equivalent to

$$w(\theta \mid \phi) = w(\theta),$$

when $w(\phi) \neq 0$.

We would hope that any sensible probability function would satisfy WIP, being that it seems to satisfy common sense and is (perhaps even overly, see [17]) weak, but it turns out that the situation is more interesting than that.

Indeed relatively few probability functions seem to satisfy WIP. We already know (from [18]) that Carnap's Continuum doesn't, for instance, except for at its end points.

We therefore ask;

"Which probability functions, satisfying Ax, also satisfy WIP?"

Thus forming our motivating question for this part of the chapter.

We do not propose to answer this question fully, rather we attempt to get us closer to the truth.

Towards this we start by looking at the case where one of our sentences contains just one constant. Here we find that WIP follows directly from Ax alone.

We then review the situation with the Paris-Nix continuum and WIP, offering a proof (corrected from [23]) that the probability functions defined by the Paris-Nix continuum (the w^δ) satisfy the principle.

In the next section we give the main result of the chapter, giving a method and proving that it generates new probability functions that satisfy WIP, when given two that already do.

We finally offer a preliminary conjecture that the only probability functions that satisfy WIP and Ax are precisely those that belong to the Paris-Nix Continuum or can be generated from them by utilising the method given previously.

We will then switch our attention to looking at the concept of *Recoverable* probability functions, the idea being that we can recover our original probability function by conditioning on future events.

Our motivation here is to find which probability functions satisfy Ax and are Recoverable.

We start by stating a couple of results which make it easier to work with the concept, by showing that we need only consider it with reference to state descriptions, and then showing that we cannot make the concept more general without avoiding triviality.

Finally we show that the probability functions which satisfy Ax, Language Invariance and are Recoverable are precisely those defined by the Paris-Nix continuum, thus providing us with an alternative characterisation of it.

4.2 The Weak Irrelevance Principle

4.2.1 Our Language

In this section we set up the language we'll be using throughout our work on WIP.

Let L_1 be the language that contains predicates $P_1(x), P_2(x), \ldots P_p(x)$, L_2 the language that contains $P_{p+1}(x), \ldots, P_{p+q}(x)$, $L = L_1 \cup L_2$.

Let $\alpha_i(x), \beta_j(x)$ list the atoms of L_1, L_2 respectively for $1 \leq i \leq 2^p$, $1 \leq j \leq 2^q$.

The atoms of L, therefore, are $\alpha_i(x) \wedge \beta_j(x)$

4.2.2 A Single Constant

In this section we aim to show that WIP follows directly from Ax whenever ϕ is defined over one constant.

Before we proceed we make everything a little easier by proving a very useful lemma.

Lemma 4.2.1. *In the statement of WIP it is sufficient to take $\theta \in SL_1, \phi \in SL_2$ to be state descriptions over their respective languages.*

Proof. Assume WIP for such state descriptions and let $\theta(a_1, a_2, \ldots, a_m) \in SL_1$, $\phi(a_{m+1}, a_{m+2}, \ldots, a_{m+k}) \in SL_2$. By the DNFT we can write

$$\theta(a_1, a_2, \ldots, a_m) = \bigvee_{s=1}^{S} \bigwedge_{i=1}^{m} \alpha_{h_{si}}(a_i),$$

for some $S \geq 1$, and

$$\phi(a_{m+1}, a_{m+2}, \ldots, a_{m+k}) = \bigvee_{t=1}^{T} \bigwedge_{j=1}^{k} \beta_{g_{tj}}(a_{m+j}),$$

for some $T \geq 1$. Then we have that:

$$\begin{aligned} w(\theta \wedge \phi) &= w\left(\bigvee_{s=1}^{S} \bigwedge_{i=1}^{m} \alpha_{h_{si}}(a_i) \wedge \bigvee_{t=1}^{T} \bigwedge_{j=1}^{k} \beta_{g_{tj}}(a_{m+j})\right) \\ &= \sum_{s=1}^{S} \sum_{t=1}^{T} w\left(\bigwedge_{i=1}^{m} \alpha_{h_{si}}(a_i) \wedge \bigwedge_{j=1}^{k} \beta_{g_{tj}}(a_{m+j})\right) \\ &= \sum_{s=1}^{S} \sum_{t=1}^{T} w\left(\bigwedge_{i=1}^{m} \alpha_{h_{si}}(a_i)\right) \cdot w\left(\bigwedge_{j=1}^{k} \beta_{g_{tj}}(a_{m+j})\right), \end{aligned}$$

because w satisfies WIP over state descriptions,

$$\begin{aligned} &= w\left(\bigvee_{s=1}^{S} \bigwedge_{i=1}^{m} \alpha_{h_{si}}(a_i)\right) \cdot w\left(\bigvee_{t=1}^{T} \bigwedge_{j=1}^{k} \beta_{g_{tj}}(a_{m+j})\right) \\ &= w(\theta) \cdot w(\phi), \end{aligned}$$

as required. □

We now proceed to the main theorem of this section

Theorem 4.2.2. *For w satisfying Ax and $\phi(a_1)$, $\theta(a_2, a_3, \ldots, a_{m+1})$ having no predicate symbols in common, we have that*

$$w(\theta(a_2, a_3, \ldots a_{m+1}) \mid \phi(a_1)) = w(\theta(a_2, a_3, \ldots a_{m+1})).$$

Proof. By Lemma 4.2.1 it is enough to consider θ, ϕ as state descriptions over L_1, L_2 respectively. Further, since ϕ is defined over one constant, it will be equivalent to an atom of L_2.

We take (without loss of generality)

$$\begin{aligned} \theta(a_2, a_3, \ldots, a_{m+1}) &= \bigwedge_{i=1}^{m} \alpha_{h_i}(a_{1+i}), \\ \phi(a_1) &= \beta_1(a_1). \end{aligned}$$

We can re-write these in terms of the atoms of L;

$$\begin{aligned} \theta(a_2, a_3, \ldots, a_{m+1}) &\equiv \bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}), \\ \phi(a_1) &\equiv \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_1(a_1), \end{aligned}$$

which means that we have

$$\theta(a_2, a_3, \ldots a_{m+1}) \wedge \phi(a_1) \equiv \bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_1(a_1). \quad (4.1)$$

We have that, by Ax,

$$w(\phi(a_1)) = \frac{1}{2^q}. \quad (4.2)$$

If we consider the permutation of atoms of L which swaps $\alpha_i(x) \wedge \beta_1(x)$ with $\alpha_i(x) \wedge \beta_t(x)$ for each $i = 1, 2, \ldots 2^p$, but leaves all other atoms fixed, then by Ax we have that

$$\begin{aligned} & w\left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_1(a_1)\right) \\ &\quad = w\left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_t(a_1)\right), \end{aligned} \quad (4.3)$$

for any $t = 1, 2, \ldots, 2^q$. We also know that

$$\bigvee_{f=1}^{2^p} \bigvee_{t=1}^{2^q} \alpha_f(x) \wedge \beta_t(x),$$

forms a tautology, so

$$\begin{aligned}
w(\theta(a_2, a_3, \ldots a_m)) &= w\left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i})\right) \\
&= w\left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \bigvee_{t=1}^{2^q} \alpha_f(a_1) \wedge \beta_t(a_1))\right) \\
&= w\left(\bigvee_{t=1}^{2^q} \left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_t(a_1)\right)\right) \\
&= \sum_{t=1}^{2^q} w\left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_t(a_1)\right) \\
&= 2^q \cdot w\left(\bigwedge_{i=1}^{m} \bigvee_{j=1}^{2^q} \alpha_{h_i}(a_{1+i}) \wedge \beta_j(a_{1+i}) \wedge \bigvee_{f=1}^{2^p} \alpha_f(a_1) \wedge \beta_1(a_1)\right),
\end{aligned}$$

by (4.3),

$$\begin{aligned}
&= 2^q \cdot w(\theta(a_2, a_3, \ldots a_m) \wedge \phi(a_1)), \quad \text{by (4.6)}, \\
&= \frac{w(\theta(a_2, a_3, \ldots a_m) \wedge \phi(a_1))}{w(\phi(a_1))}, \quad \text{by (4.2)}, \\
&= w(\theta(a_2, a_3, \ldots, a_m) \mid \phi(a_1)),
\end{aligned}$$

as required. □

Note that, by symmetry, the result will also hold for when θ has one constant.

It remains for us to show that we cannot improve upon this result. We take our counter-example from [18], where it was used as the counter-example to w_λ satisfying WIP, and where both θ and ϕ are defined over two constants.

Let

$$L_1 = \{P_1\}, L_2 = \{P_2\},$$

$$\theta(a_3, a_4) = P_2(a_3) \wedge P_2(a_4),$$

$$\phi(a_1, a_2) = P_1(a_1) \wedge P_1(a_2).$$

Then, by the definition of w_λ, given on page 25, we have that

$$\begin{aligned} w_\lambda(\phi(a_1, a_2)) &= \frac{2(1+\frac{\lambda}{4})\frac{\lambda}{4} + 2(\frac{\lambda}{4}^2)}{\lambda(1+\lambda)} \\ &= \frac{2+\lambda/2+\lambda/2}{4(1+\lambda)} \\ &= \frac{2+\lambda}{4(1+\lambda)}, \end{aligned}$$

$$\begin{aligned} w_\lambda(\theta(a_3, a_4) \wedge \phi(a_1, a_2)) &= \frac{(3+\frac{\lambda}{4})(2+\frac{\lambda}{4})(1+\frac{\lambda}{4})\frac{\lambda}{4} + 4(2+\frac{\lambda}{4})(1+\frac{\lambda}{4})(\frac{\lambda}{4})^2}{\lambda(1+\lambda)(2+\lambda)(3+\lambda)} + \\ &\quad \frac{3(1+\frac{\lambda}{4})^2(\frac{\lambda}{4})^2 + 8(1+\frac{\lambda}{4})(\frac{\lambda}{4})^3}{\lambda(1+\lambda)(2+\lambda)(3+\lambda)} \\ &= \frac{(1+\frac{\lambda}{4})((3+\frac{\lambda}{4})(2+\frac{\lambda}{4}) + 4(2+\frac{\lambda}{4})(\frac{\lambda}{4}) + 3(1+\frac{\lambda}{4})(\frac{\lambda}{4}) + 2(\frac{\lambda^2}{4}))}{4(1+\lambda)(2+\lambda)(3+\lambda)} \\ &= \frac{(4+\lambda)(6+4\lambda+\lambda^2)}{16(1+\lambda)(2+\lambda)(3+\lambda)}. \end{aligned}$$

So

$$\begin{aligned} w_\lambda(\theta(a_3, a_4) \mid \phi(a_1, a_2)) &= \frac{w(\theta(a_3, a_4) \wedge \phi(a_1, a_2))}{w(\theta(a_3, a_4))} \\ &= \frac{4(1+\lambda)(4+\lambda)(6+4\lambda+\lambda^2)}{16(2+\lambda)(1+\lambda)(2+\lambda)(3+\lambda)} \\ &= \frac{(4+\lambda)(6+4\lambda+\lambda^2)}{4(2+\lambda)^2(3+\lambda)}. \end{aligned}$$

Our aim now is to show that $w(\theta(a_3, a_4) \mid \phi(a_1, a_2)) \neq w(\theta(a_3, a_4))$. Specifically, we're going to show that $w(\theta(a_3, a_4) \mid \phi(a_1, a_2)) > w(\theta(a_3, a_4))$;

$$\begin{aligned} w_\lambda(\theta(a_3, a_4) \mid \phi(a_1, a_2)) &> w(\theta(a_3, a_4)) \\ \iff \frac{(4+\lambda)(6+4\lambda+\lambda^2)}{4(2+\lambda)^2(3+\lambda)} &> \frac{2+\lambda}{4(1+\lambda)} \\ \iff (1+\lambda)(4+\lambda)(6+4\lambda+\lambda^2) &> (2+\lambda)^3(3+\lambda) \\ \iff 24+46\lambda+30\lambda^2+9\lambda^3+\lambda^4 &> 24+44\lambda+30\lambda^2+9\lambda^3+\lambda^4, \end{aligned}$$

which, since $\lambda > 0$, gives us our counter-example as required.

4.2.3 WIP and the Paris-Nix Continuum

In [23] a proof is offered that the Paris-Nix continuum satisfies WIP. Unfortunately this proof contains errors (the most serious of which cancel each other out, giving the appearance that the proof is correct), and does not include the special case of $\delta = 1$. Here we attempt to provide a corrected, completed proof.

Theorem 4.2.3. *If w belongs to the Paris-Nix Continuum, then it satisfies WIP.*

Proof. Let L_1, L_2, and L be as before, $\theta(a_1, a_2, \ldots, a_m) \in SL_1$, $\phi(a_{m+1}, \ldots, a_{m+k}) \in SL_2$. By Lemma 4.2.1 it is enough to take θ, ϕ to be state descriptions. Therefore we take, without loss of generality,

$$\theta(a_1, a_2, \ldots, a_m) = \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i),$$

$$\phi(a_{m+1}, \ldots, a_{m+k}) = \bigwedge_{j=1}^{k} \beta_{g_j}(a_{m+j}).$$

Let $m_r = |\{i \mid 1 \leq i \leq m, h_i = r\}|$, $r = 1, \ldots, 2^p$, $k_s = |\{j \mid 1 \leq j \leq k, g_j = s\}|$, $s = 1, \ldots, 2^q$.

In L we have that

$$\theta(a_1, a_2, \ldots, a_m) \equiv \bigvee_{1 \leq t_1, \ldots, t_m \leq 2^q} \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \wedge \beta_{t_i}(a_i), \tag{4.4}$$

$$\phi(a_{m+1}, \ldots, a_{m+k}) \equiv \bigvee_{1 \leq f_1, \ldots, f_k \leq 2^p} \bigwedge_{j=1}^{k} \alpha_{f_j}(a_{m+j}) \wedge \beta_{g_j}(a_{m+j}).$$

We are initially going to concentrate on $w(\theta)$.

For any $u \geq 0$ we have that the atom $\alpha_r(x) \wedge \beta_s(x)$ occurs u times in the disjunct $\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \wedge \beta_{t_i}(a_i)$ for exactly

$$\binom{m_r}{u} (2^q - 1)^{(m_r - u)} 2^{q(m - m_r)}$$

choices of $1 \leq t_1, t_2, \ldots t_m \leq 2^q$.

In the sequel we will be using the following definition of w^δ (adapted from [23], p.29), which holds for $\delta \in [0, 1)$[1].

$$w^\delta \left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \wedge \beta_{t_i}(a_i) \right) = \frac{\gamma^m}{2^{p+q}} \sum_{\substack{1 \leq r \leq 2^p \\ 1 \leq s \leq 2^q}} \left(1 + \frac{\delta}{\gamma} \right)^{m_{rs}},$$

[1] We later address the special case where $\delta = 1$.

where $2^n\gamma = 1-\delta$, $m_{rs} = |\{i : 1 \le i \le m,\ r = h_i,\ s = t_i\}|$.

Then we have that

$$\begin{aligned}
w(\theta) &= w\left(\bigvee_{1\le t_1,\dots,t_m\le 2^q}\bigwedge_{i=1}^{m}\alpha_{h_i}(a_i)\wedge\beta_{t_i}(a_i)\right)\\
&= \frac{\gamma^m}{2^{p+q}}\sum_{r=1}^{2^p}\sum_{s=1}^{2^q}\sum_{u=0}^{m_r}\binom{m_r}{u}(2^q-1)^{m_r-u}\left(1+\frac{\delta}{\gamma}\right)^u 2^{q(m-m_r)}\\
&= \frac{\gamma^m}{2^{p+q}}\sum_{r=1}^{2^p}\sum_{s=1}^{2^q}\left(2^q+\frac{\delta}{\gamma}\right)^{m_r}2^{q(m-m_r)}\\
&= \frac{\gamma^m}{2^{p+q}}\sum_{r=1}^{2^p}\left(2^q+\frac{\delta}{\gamma}\right)^{m_r}2^{q(m-m_r)}2^q
\end{aligned}$$

and analogously

$$w(\phi) = \frac{\gamma^k}{2^{p+q}}\sum_{s=1}^{2^q}\left(2^p+\frac{\delta}{\gamma}\right)^{k_s}2^{p(k-k_s)}2^p,$$

so,

$$\begin{aligned}
w(\theta)\cdot w(\phi) &= \left(\frac{\gamma^m}{2^{p+q}}\sum_{r=1}^{2^p}\left(2^q+\frac{\delta}{\gamma}\right)^{m_r}2^{q(m-m_r)}2^q\right)\cdot\left(\frac{\gamma^k}{2^{p+q}}\sum_{s=1}^{2^q}\left(2^p+\frac{\delta}{\gamma}\right)^{k_s}2^{p(k-k_s)}2^p\right)\\
&= \frac{\gamma^{m+k}}{2^{2(p+q)}}\sum_{r=1}^{2^p}\sum_{s=1}^{2^q}\left(2^q+\frac{\delta}{\gamma}\right)^{m_r}2^{q(m-m_r)}2^q\left(2^p+\frac{\delta}{\gamma}\right)^{k_s}2^{p(k-k_s)}2^p\\
&= \frac{\gamma^{m+k}}{2^{2(p+q)}}2^p2^q\sum_{r=1}^{2^p}\sum_{s=1}^{2^q}\left(2^q+\frac{\delta}{\gamma}\right)^{m_r}\left(2^p+\frac{\delta}{\gamma}\right)^{k_s}2^{q(m-m_r)}2^{p(k-k_s)}\\
&= \frac{\gamma^{m+k}}{2^{p+q}}\sum_{r=1}^{2^p}\sum_{s=1}^{2^q}\left(2^q+\frac{\delta}{\gamma}\right)^{m_r}\left(2^p+\frac{\delta}{\gamma}\right)^{k_s}2^{q(m-m_r)}2^{p(k-k_s)}. \qquad (4.5)
\end{aligned}$$

Now we consider the sentence $w(\theta(a_1,a_2,\dots a_m)\wedge\phi(a_{m+1},\dots,a_{m+q}))$. If we consider the atoms of L, then we can write the sentence as

$$\bigvee_{1\le t_1,\dots,t_m\le 2^q}\ \bigvee_{1\le f_1,\dots,f_k\le 2^p}\ \bigwedge_{i=1}^{m}\alpha_{h_i}(a_i)\wedge\beta_{t_i}(a_i)\wedge\bigwedge_{j=1}^{k}\alpha_{f_j}(a_{m+j})\wedge\beta_{g_j}(a_{m+j}). \qquad (4.6)$$

For any $u\ge 0$, we have that the atom $\alpha_r(x)\wedge\beta_s(x)$ occurs u times in the disjunct $\bigwedge_{i=1}^{m}\alpha_{h_i}(a_i)\wedge\beta_{t_i}(a_i)\wedge\bigwedge_{j=1}^{k}\alpha_{f_j}(a_{m+j})\wedge\beta_{g_j}(a_{m+j})$ for exactly

$$\sum_{\substack{0\le x\le m_r\\0\le y\le k_s\\x+y=u}}\binom{m_r}{x}(2^q-1)^{m_r-x}2^{q(m-m_r)}\binom{k_s}{y}(2^p-1)^{k_s-y}2^{p(k-k_s)} \qquad (4.7)$$

choices of $1 \leq t_1, \ldots, t_m \leq 2^q$ and $1 \leq f_1, \ldots f_k \leq 2^p$. Therefore

$$\begin{aligned}
w(\theta \wedge \phi) &= w\left(\bigvee_{1\leq t_1,\ldots,t_m\leq 2^q}\ \bigvee_{1\leq f_1,\ldots,f_k\leq 2^p}\ \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \wedge \beta_{t_i}(a_i) \wedge \bigwedge_{j=1}^{k} \alpha_{f_j}(a_{m+j}) \wedge \beta_{g_j}(a_{m+j})\right) \\
&= \frac{\gamma^{m+k}}{2^{p+q}} \sum_{r=1}^{2^p} \sum_{s=1}^{2^q} \sum_{u=0}^{m_r+k_s} \sum_{\substack{0\leq x\leq m_r \\ 0\leq y\leq k_s \\ x+y=u}} \\
&\qquad \binom{m_r}{x} (2^q-1)^{m_r-x} 2^{q(m-m_r)} \binom{k_s}{y} (2^p-1)^{k_s-y} 2^{p(k-k_s)} \left(1+\frac{\delta}{\gamma}\right)^u \\
&= \frac{\gamma^{m+k}}{2^{p+q}} \sum_{r=1}^{2^p} \sum_{s=1}^{2^q} \left(2^q+\frac{\delta}{\gamma}\right)^{m_r} \left(2^p+\frac{\delta}{\gamma}\right)^{k_s} 2^{q(m-m_r)} 2^{p(k-k_s)} \\
&= w(\theta) \cdot w(\phi),
\end{aligned}$$

by (4.5), as required.

Now we consider the special case of $\delta = 1$, where w^δ collapses to Carnap's $\mathfrak{m}_0$. We start by looking at $w(\theta)$.

If $m_r < m$ for $1 \leq r \leq 2^p$, then the definition of θ given by (4.4) cannot possibly include a homogenous state description since every state description involved in the disjunct has at least two different α_{h_i}'s. Therefore, by the definition of $\mathfrak{m}_0$, $w(\theta) = 0$.

Similarly every state description in the disjunct of (4.6) must contain at least two different α_{h_i}'s, so $w(\theta \wedge \phi) = 0$.

Therefore

$$w(\theta \wedge \phi) = 0 = w(\theta) \cdot w(\phi),$$

as required. Also, by symmetry the same result holds for $w(\phi)$ if $k_s < k$ for $1 \leq s \leq 2^q$.

We now take the case where there exists some m_r such that $m_r = m$ and some k_s such that $k_s = k$.

Again, we start by considering $w(\theta)$.

For each $1 \leq t \leq 2^q$ we will have one state description in the disjunct $\bigwedge_{i=1}^m \alpha_{h_i}(a_i) \wedge \beta_{t_i}(a_i)$ that forms a homogenous state description, namely the one where we take

$$t_1 = t_2 = \ldots = t_i = \ldots = t_m = t$$

Each of these 2^q homogenous state descriptions contribute $\frac{1}{2^{p+q}}$ to $w(\theta)$ by the definition of $\mathfrak{m}_0$.

All other choices for $1 \leq t_1, t_2, \ldots, t_m \leq 2^q$ produce non-homogenous state descriptions, each of which contribute 0 to the probability of θ. Therefore[2]

$$w(\theta) = 2^q \cdot \frac{1}{2^{p+q}} = \frac{1}{2^p},$$

and similarly,

$$w(\phi) = 2^p \cdot \frac{1}{2^{p+q}} = \frac{1}{2^q}.$$

We now consider $w(\theta \wedge \phi)$.

Here we can only consider one homogenous state description, that based upon the atom $\alpha_r(x) \wedge \beta_s(x)$.

By (4.7) there is one occurrence of this state description, with probability $\frac{1}{2^{p+q}}$. Therefore we have that :

$$\begin{aligned} w(\theta \wedge \phi) &= \frac{1}{2^{p+q}} \\ &= \frac{1}{2^p} \cdot \frac{1}{2^q} \\ &= w(\theta) \cdot w(\phi), \end{aligned}$$

as required. $\square$

[2] We could have done this by staying in L_1, however we choose to give the calculation in L because it makes it easier to see where $w(\theta \wedge \phi)$ comes from.

4.2.4 Generating Further Probability Functions That Satisfy WIP

In this section we show how we can generate further probability functions that satisfy WIP, by starting with two that already do.

Our language is as defined in Section 4.2.1.

Let w_1, w_2 be discrete probability functions whose de Finetti priors put measure e_x on point $\vec{c_x} \in \mathbb{D}_{2^{p+q}}$ for $x = 1, 2, \ldots, X$ and measure f_y on point $\vec{b_y} \in \mathbb{D}_{2^{p+q}}$ for $y = 1, 2, \ldots, Y$ respectively. Let $\lambda \in [0, 1]$.

Define the discrete probability function, $w = w_1 \oplus_\lambda w_2$ to put measure

$$\sum \{e_x f_y \mid \vec{k} = \lambda \vec{c_x} + (1-\lambda)\vec{b_y}, 1 \leq x \leq X, 1 \leq y \leq Y\},$$

on point $\vec{k} \in \mathbb{D}_{2^{p+q}}$.

Note that this means if the $\lambda \vec{c_x} + (1-\lambda)\vec{b_y}$ are distinct then w just puts measure $e_x f_y$ on this point.

Theorem 4.2.4. *If w_1, w_2 are discrete probability functions on L satisfying Ex, Ax, Language Invariance and WIP and $\lambda \in [0, 1]$ then $w = w_1 \oplus_\lambda w_2$ also satisfies these principles.*

Proof. Let w_1', w_2' be the Language Invariant forms of w_1, w_2 defined on language L_1 by the de Finetti priors that put measure e_x', on points $\vec{c_x'} \in \mathbb{D}_{2^p}$ for $x = 1, 2, \ldots X'$ and measure f_y' on point $\vec{b_y'} \in \mathbb{D}_{2^p}$ for $y = 1, 2, \ldots Y'$, respectively. Let w_1'', w_2'' be similarly defined for L_2.

Since the w, w_1, w_2 are each defined by a discrete measure, the result is trivial for Ex. We consider each of the other principles in turn.

Ax

We take θ to be some state description of[3] L_1 defined over m constants. Without loss of generality define it to have the vector representation given by

$$\langle m_1, m_2, \ldots, m_r, \ldots, m_{2^p} \rangle.$$

For the purposes of simplification of notation, we assume that the coordinate of $\vec{c'_x}$ corresponding to the atom α_r is c'_{xr}, and similarly for $\vec{b'_y}$. The probability given to ψ by the function w' is therefore

$$w'(\theta) = \sum_{x=1}^{X'} \sum_{y=1}^{Y'} e'_x f'_y \prod_{r=1}^{2^p} \left(\lambda c'_{xr} + (1-\lambda) b'_{yr}\right)^{m_r}. \tag{4.8}$$

Let σ be any permutation of $\{1, 2, \ldots, 2^p\}$ and define

$$\sigma(\theta) = \alpha_{\sigma(1)}^{m_1} \wedge \alpha_{\sigma(2)}^{m_2} \wedge \ldots \wedge \alpha_{\sigma(2^p)}^{m_{2^p}}.$$

Let

$$Z_{\vec{m}} = \{\langle g_1, \ldots, g_r \rangle \mid g_t \leq m_t \text{ for } t = 1, \ldots, 2^p\},$$

then we have that

$$\begin{aligned}
w'(\sigma(\theta)) &= \sum_{x=1}^{X'} \sum_{y=1}^{Y'} e'_x f'_y \prod_{r=1}^{2^p} \left(\lambda c'_{x\sigma(r)} + (1-\lambda) b'_{y\sigma(r)}\right)^{m_r} \\
&= \sum_{x=1}^{X'} \sum_{y=1}^{Y'} e'_x f'_y \prod_{r=1}^{2^p} \left(\sum_{g_r=0}^{m_r} \binom{m_r}{g_r} \lambda^{g_r} (1-\lambda)^{m_r-g_r} c'^{g_r}_{x\sigma(r)} b'^{m_r-g_r}_{y\sigma(r)} \right) \\
&= \sum_{x=1}^{X'} \sum_{y=1}^{Y'} e'_x f'_y \left(\sum_{\langle g_1,\ldots,g_r \rangle \in Z_{\vec{m}}} \prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r} (1-\lambda)^{m_r-g_r} c'^{g_r}_{x\sigma(r)} b'^{m_r-g_r}_{y\sigma(r)} \right) \\
&= \sum_{\langle g_1,\ldots,g_r \rangle \in Z_{\vec{m}}} \left(\sum_{y=1}^{Y'} f'_y \prod_{r=1}^{2^p} b'^{m_r-g_r}_{y\sigma(r)} \right) \left(\sum_{x=1}^{X'} e'_x \prod_{r=1}^{2^p} c'^{g_r}_{x\sigma(r)} \right) \\
&\qquad \left(\prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r} (1-\lambda)^{m_r-g_r} \right) \\
&= \sum_{\langle g_1,\ldots,g_r \rangle \in Z_{\vec{m}}} \left(\sum_{y=1}^{Y'} f'_y \prod_{r=1}^{2^p} b'^{m_r-g_r}_{yr} \right) \left(\sum_{x=1}^{X'} e'_x \prod_{r=1}^{2^p} c'^{g_r}_{xr} \right) \\
&\qquad \left(\prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r} (1-\lambda)^{m_r-g_r} \right),
\end{aligned}$$

[3]Here we use L_1 to simplify our notation, the result clearly carries through for any language.

by Ax for w'_1, w'_2,

$$\begin{aligned}
&= \sum_{x=1}^{X'}\sum_{y=1}^{Y'} e'_x f'_y \sum_{\langle g_1,\ldots,g_r\rangle \in Z_{\vec{m}}} \left(\prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r}(1-\lambda)^{m_r-g_r} c'^{g_r}_{xr} b'^{m_r-g_r}_{yr} \right) \\
&= \sum_{x=1}^{X'}\sum_{y=1}^{Y'} e'_x f'_y \left(\prod_{r=1}^{2^p} \lambda c'_{xr} + (1-\lambda) b'_{yr} \right)^{m_r} \\
&= w'(\theta),
\end{aligned}$$

as required. Therefore w' satisfies Ax whenever w'_1, w'_2 do.

LIP

Let θ be as defined as above. The corresponding definition of θ over L is

$$\theta = \bigwedge_{r=1}^{2^p} \left(\bigvee_{t=1}^{2^q} \alpha_r \wedge \beta_t \right)^{m_r}.$$

Throughout the following, for the purposes of simplification of notation, we assume that the coordinate of $\vec{c_x}$ corresponding to the atom $\alpha_r \wedge \beta_s$ is $c_{x(rs)}$, and similarly for $\vec{b_y}$. The probability of θ under w is therefore

$$w(\theta) = \sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \prod_{r=1}^{2^p} \left(\lambda \left(\sum_{s=1}^{2^q} c_{x(rs)} \right) + (1+\lambda) \left(\sum_{s=1}^{2^q} b_{y(rs)} \right) \right)^{m_r}$$

We need to show that this is the same as the probability $w'(\theta)$, given in (4.8). We proceed in a similar fashion as for our proof for Ax;

$$\begin{aligned}
w(\theta) &= \sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \prod_{r=1}^{2^p} \left(\lambda \left(\sum_{s=1}^{2^q} c_{x(rs)} \right) + (1-\lambda) \left(\sum_{s=1}^{2^q} b_{y(rs)} \right) \right)^{m_r} \\
&= \sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \Bigg(\sum_{\langle g_1,\ldots,g_r\rangle \in Z_{\vec{m}}} \prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r}(1-\lambda)^{m_r-g_r} \\
&\qquad \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r-g_r} \Bigg) \\
&= \sum_{\langle g_1,\ldots g_r\rangle \in Z_{\vec{m}}} \left(\sum_{y=1}^{Y'} f'_y \prod_{r=1}^{2^p} b'^{m_r-g_r}_{yr} \right) \left(\sum_{x=1}^{X'} e'_x \prod_{r=1}^{2^p} c'^{g_r}_{xr} \right) \\
&\qquad \left(\prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r}(1-\lambda)^{m_r-g_r} \right),
\end{aligned}$$

because w_1', w_2' are the language invariant versions of w_1, w_2 over L_1,

$$\begin{aligned}
&= \sum_{x=1}^{X'}\sum_{y=1}^{Y'} e_x' f_y' \left(\sum_{\langle g_1,\ldots,g_r\rangle \in Z_{\vec{m}}} \prod_{r=1}^{2^p} \binom{m_r}{g_r} \lambda^{g_r}(1-\lambda)^{m_r-g_r} c_{xr}'^{g_r} b_{yr}'^{m_r-g_r} \right)\\
&= \sum_{x=1}^{X'}\sum_{y=1}^{Y'} e_x' f_y' \left(\prod_{r=1}^{2^p} \lambda c_{xr}' + (1-\lambda) b_{yr}' \right)^{m_r}\\
&= w'(\theta),
\end{aligned}$$

as required. Therefore whenever w_1, w_2 are Language Invariant, so too is w.

WIP

By Lemma 4.2.1 it is enough to take θ, ϕ to be state descriptions. Define θ to be some state description of the language L_1 and define ϕ to be some state description of the language L_2. We aim to show that for all such state descriptions

$$\begin{aligned}
&w(\theta(a_1, a_2, \ldots, a_m) \wedge \phi(a_{m+1}, \ldots, a_{m+k}))\\
&\qquad = w(\theta(a_1, a_2, \ldots, a_m)) \cdot w(\phi(a_{m+1}, \ldots, a_{m+k})), \qquad (4.9)
\end{aligned}$$

given that we know

$$\begin{aligned}
&w_1(\theta(a_1, a_2, \ldots, a_m) \wedge \phi(a_{m+1}, \ldots, a_{m+k}))\\
&\qquad = w_1(\theta(a_1, a_2, \ldots, a_m)) \cdot w_1(\phi(a_{m+1}, \ldots, a_{m+k})), \qquad (4.10)
\end{aligned}$$

and

$$\begin{aligned}
&w_2(\theta(a_1, a_2, \ldots, a_m) \wedge \phi(a_{m+1}, \ldots, a_{m+k}))\\
&\qquad = w_2(\theta(a_1, a_2, \ldots, a_m)) \cdot w_2(\phi(a_{m+1}, \ldots, a_{m+k})). \qquad (4.11)
\end{aligned}$$

We take $\theta \in SL_1$ to be defined as

$$\theta = \bigwedge_{r=1}^{2^p} (\alpha_r)^{m_r},$$

with its corresponding definition over L being

$$\theta = \bigwedge_{r=1}^{2^p} \left(\bigvee_{t=1}^{2^q} \alpha_r \wedge \beta_t \right)^{m_r}.$$

Similarly define $\phi \in SL_2$ as

$$\phi = \bigwedge_{g=1}^{2^q} (\beta_g)^{k_s},$$

with its corresponding definition over L being

$$\phi = \bigwedge_{g=1}^{2^q} \left(\bigvee_{f=1}^{2^p} \alpha_f \wedge \beta_g \right)^{k_s}.$$

We can now re-write (4.9) as follows;

$$\begin{aligned} & w\left(\bigwedge_{r=1}^{2^p} \left(\bigvee_{t=1}^{2^q} \alpha_r \wedge \beta_t \right)^{m_r} \wedge \bigwedge_{g=1}^{2^q} \left(\bigvee_{f=1}^{2^p} \alpha_f \wedge \beta_g \right)^{k_s} \right) \\ & \quad = w\left(\bigwedge_{r=1}^{2^p} \left(\bigvee_{t=1}^{2^q} \alpha_r \wedge \beta_t \right)^{m_r} \right) \cdot w\left(\bigwedge_{g=1}^{2^q} \left(\bigvee_{f=1}^{2^p} \alpha_f \wedge \beta_g \right)^{k_s} \right), \end{aligned}$$

and similarly (4.10) & (4.11).

With the previous conventions for $c_{x(rs)}$, $b_{y(rs)}$ in mind, we can write what we want to show in terms of the actual probability of the sentences as follows

$$\begin{aligned} & \sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \prod_{r=1}^{2^p} \left(\lambda \left(\sum_{s=1}^{2^q} c_{x(rs)} \right) + (1-\lambda) \left(\sum_{s=1}^{2^q} b_{y(rs)} \right) \right)^{m_r} \\ & \qquad \prod_{d=1}^{2^q} \left(\lambda \left(\sum_{h=1}^{2^p} c_{x(hd)} \right) + (1-\lambda) \left(\sum_{h=1}^{2^p} b_{y(hd)} \right) \right)^{k_d} \\ & = \left(\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \prod_{r=1}^{2^p} \left(\lambda \left(\sum_{s=1}^{2^q} c_{x(rs)} \right) + (1-\lambda) \left(\sum_{s=1}^{2^q} b_{y(rs)} \right) \right)^{m_r} \right) \\ & \qquad \times \left(\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \prod_{d=1}^{2^q} \left(\lambda \left(\sum_{h=1}^{2^p} c_{x(hd)} \right) + (1-\lambda) \left(\sum_{h=1}^{2^p} b_{y(hd)} \right) \right)^{k_d} \right), \end{aligned} \tag{4.12}$$

given that we know

$$\begin{aligned} & \sum_{x=1}^{X} e_x \left(\prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{m_r} \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{k_d} \right) \\ & \quad = \left(\sum_{x=1}^{X} e_x \prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{m_r} \right) \times \left(\sum_{x=1}^{X} e_x \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{k_d} \right), \end{aligned} \tag{4.13}$$

and

$$\sum_{y=1}^{Y} f_y \left(\prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r} \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d} \right)$$
$$= \left(\sum_{y=1}^{Y} f_y \prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r} \right) \times \left(\sum_{y=1}^{Y} f_y \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d} \right). \quad (4.14)$$

If we expand the LHS of (4.12), setting

$$\begin{aligned} Z_{\vec{m}} &= \{\langle g_1, \ldots, g_r \rangle \mid g_t \leq m_t \text{ for } t = 1, \ldots, 2^p\}, \\ Z_{\vec{k}} &= \{\langle v_1, \ldots, v_d \rangle \mid v_u \leq k_u \text{ for } u = 1, \ldots, 2^q\}, \end{aligned}$$

we get

$$\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \left(\prod_{r=1}^{2^p} \left(\sum_{g_r=0}^{m_r} \binom{m_r}{g_r} \lambda^{g_r} (1-\lambda)^{m_r - g_r} \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r - g_r} \right) \right.$$
$$\left. \prod_{d=1}^{2^q} \left(\sum_{v_d=0}^{k_d} \binom{k_d}{v_d} \lambda^{v_d} (1-\lambda)^{k_d - v_d} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{v_d} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d - v_d} \right) \right),$$

multiplying this out gives

$$\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \left(\sum_{\substack{\langle g_1, \ldots, g_r \rangle \in Z_{\vec{m}} \\ \langle v_1, \ldots, v_d \rangle \in Z_{\vec{k}}}} \prod_{\substack{1 \leq r \leq 2^p \\ 1 \leq d \leq 2^q}} \left(\binom{m_r}{g_r} \binom{k_d}{v_d} \lambda^{g_r + v_d} (1-\lambda)^{m_r + k_d - g_r - v_d} \right) \right.$$
$$\left. \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r - g_r} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{v_d} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d - v_d} \right) \Bigg).$$

Re-arranging gives

$$\sum_{\substack{\langle g_1, \ldots, g_r \rangle \in Z_{\vec{m}} \\ \langle v_1, \ldots, v_d \rangle \in Z_{\vec{k}}}} \prod_{\substack{1 \leq r \leq 2^p \\ 1 \leq d \leq 2^q}} \left(\binom{m_r}{g_r} \binom{k_d}{v_d} \lambda^{g_r + v_d} (1-\lambda)^{m_r + k_d - g_r - v_d} \right)$$
$$\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \left(\prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r - g_r} \right.$$
$$\left. \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{v_d} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d - v_d} \right),$$

and separating the x and y parts gives

$$\sum_{\substack{\langle g_1,\ldots,g_r\rangle\in Z_{\vec{m}}\\ \langle v_1,\ldots,v_d\rangle\in Z_{\vec{k}}}} \left(\prod_{\substack{1\le r\le 2^p\\ 1\le d\le 2^q}} \binom{m_r}{g_r}\binom{k_d}{v_d} \lambda^{g_r+v_d}(1-\lambda)^{m_r+k_d-g_r-v_d} \right)$$
$$\left(\sum_{x=1}^{X} e_x \prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{v_d} \right)$$
$$\times \left(\sum_{y=1}^{Y} f_y \prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r-g_r} \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d-v_d} \right).$$

By applying (4.13) and (4.14) we obtain

$$\sum_{\substack{\langle g_1,\ldots,g_r\rangle\in Z_{\vec{m}}\\ \langle v_1,\ldots,v_d\rangle\in Z_{\vec{k}}}} \left(\prod_{\substack{1\le r\le 2^p\\ 1\le d\le 2^q}} \binom{m_r}{g_r}\binom{k_d}{v_d} \lambda^{g_r+v_d}(1-\lambda)^{m_r+k_d-g_r-v_d} \right)$$
$$\left(\left(\sum_{x=1}^{X} e_x \prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \right) \times \left(\sum_{x=1}^{X} e_x \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{v_d} \right) \right)$$
$$\times \left(\left(\sum_{y=1}^{Y} f_y \prod_{r=1}^{2^p} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r-g_r} \right)\right.$$
$$\left.\times \left(\sum_{y=1}^{Y} f_y \prod_{d=1}^{2^q} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d-v_d} \right) \right)$$
$$=$$

$$\sum_{\substack{\langle g_1,\ldots,g_r\rangle\in Z_{\vec{m}}\\ \langle v_1,\ldots,v_d\rangle\in Z_{\vec{k}}}} \prod_{\substack{1\le r\le 2^p\\ 1\le d\le 2^q}} \left(\binom{m_r}{g_r}\binom{k_d}{v_d} \lambda^{g_r+v_d}(1-\lambda)^{m_r+k_d-g_r-v_d} \right)$$
$$\left(\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \prod_{r=1}^{2^p} \left(\left(\sum_{s=1}^{2^q} c_{x(rs)} \right)^{g_r} \left(\sum_{s=1}^{2^q} b_{y(rs)} \right)^{m_r-g_r} \right) \right)$$
$$\times \left(\sum_{x=1}^{X} \sum_{y=1}^{Y} e_x f_y \prod_{d=1}^{2^q} \left(\left(\sum_{h=1}^{2^p} c_{x(hd)} \right)^{v_d} \left(\sum_{h=1}^{2^p} b_{y(hd)} \right)^{k_d-v_d} \right) \right)$$
$$=$$

$$\left(\sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \sum_{\langle g_1,\ldots,g_r\rangle \in Z_{\vec{m}}} \prod_{1\le r\le 2^p} \binom{m_r}{g_r} \lambda^{g_r}(1-\lambda)^{m_r-g_r} \left(\sum_{s=1}^{2^q} c_{x(rs)}\right)^{g_r} \left(\sum_{s=1}^{2^q} b_{y(rs)}\right)^{m_r-g_r}\right)$$

$$\times \left(\sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \sum_{\langle v_1,\ldots,v_d\rangle \in Z_{\vec{k}}} \prod_{1\le d\le 2^q} \binom{k_d}{v_d} \lambda^{v_d}(1-\lambda)^{k_d-v_d} \left(\sum_{h=1}^{2^p} c_{x(hd)}\right)^{v_d} \left(\sum_{h=1}^{2^p} b_{y(hd)}\right)^{k_d-v_d}\right)$$

$$=$$

$$\left(\sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \prod_{r=1}^{2^p} \left(\lambda\left(\sum_{s=1}^{2^q} c_{x(rs)}\right) + (1-\lambda)\left(\sum_{s=1}^{2^q} b_{y(rs)}\right)\right)^{m_r}\right)$$

$$\times \left(\sum_{x=1}^{X}\sum_{y=1}^{Y} e_x f_y \prod_{d=1}^{2^q} \left(\lambda\left(\sum_{h=1}^{2^p} c_{x(hd)}\right) + (1-\lambda)\left(\sum_{h=1}^{2^p} b_{y(hd)}\right)\right)^{k_d}\right)$$

$$=$$

$$RHS \text{ of } (4.12),$$

as required. □

This proof generalises to the continuous[4] case, as follows.

Theorem 4.2.5. *If w_1, w_2 are continuous probability functions with de Finetti priors μ_1, μ_2 respectively satisfying Ex, Ax, Language Invariance and WIP and $\lambda \in [0,1]$ then w, with de Finetti prior given by*

$$\int f(\vec{z})\, d\mu(\vec{z}) \;=\; \int\int f(\lambda\vec{x} + (1-\lambda)\vec{y})\, d\mu_1(\vec{x})\, d\mu_2(\vec{y}),$$

also satisfies these principles.

[4] We do not know if there exist any genuinely continuous probability functions that satisfy WIP, so this case may be moot.

Proof. We will just show this for WIP. In the continuous case we want to show that

$$\int\int\prod_{r=1}^{2^p}\left(\lambda\left(\sum_{s=1}^{2^q}x_{rs}\right)+(1-\lambda)\left(\sum_{s=1}^{2^q}y_{rs}\right)\right)^{m_r}$$
$$\times\prod_{d=1}^{2^q}\left(\lambda\left(\sum_{h=1}^{2^p}x_{hd}\right)+(1-\lambda)\left(\sum_{h=1}^{2^p}y_{hd}\right)\right)^{k_d}\,d\mu_1\,d\mu_2$$
$$=\int\int\prod_{r=1}^{2^p}\left(\lambda\left(\sum_{s=1}^{2^q}x_{rs}\right)+(1-\lambda)\left(\sum_{s=1}^{2^q}y_{rs}\right)\right)^{m_r}\,d\mu_1\,d\mu_2$$
$$\times\int\int\prod_{d=1}^{2^q}\left(\lambda\left(\sum_{h=1}^{2^p}x_{hd}\right)+(1-\lambda)\left(\sum_{h=1}^{2^p}y_{hd}\right)\right)^{k_d}\,d\mu_1\,d\mu_2,$$

given that we know

$$\int\prod_{r=1}^{2^p}\left(\sum_{s=1}^{2^q}x_{rs}\right)^{m_r}\prod_{d=1}^{2^q}\left(\sum_{h=1}^{2^p}x_{hd}\right)^{k_d}\,d\mu_1$$
$$=\int\prod_{r=1}^{2^p}\left(\sum_{s=1}^{2^q}x_{rs}\right)^{m_r}\,d\mu_1\times\int\prod_{d=1}^{2^q}\left(\sum_{h=1}^{2^p}x_{hd}\right)^{k_d}\,d\mu_1,$$

and

$$\int\prod_{r=1}^{2^p}\left(\sum_{s=1}^{2^q}y_{rs}\right)^{m_r}\prod_{d=1}^{2^q}\left(\sum_{h=1}^{2^p}y_{hd}\right)^{k_d}\,d\mu_2$$
$$=\int\prod_{r=1}^{2^p}\left(\sum_{s=1}^{2^q}y_{rs}\right)^{m_r}\,d\mu_2\times\int\prod_{d=1}^{2^q}\left(\sum_{h=1}^{2^p}y_{hd}\right)^{k_d}\,d\mu_2.$$

The proof now follows similarly to that for Theorem 4.2.4. □

At the moment we are aware of no other probability functions that satisfy these principles, therefore we put forward a preliminary conjecture:

Conjecture 4.2.6. *The only probability functions satisfying WIP, Language Invariance, Ex and Ax are the w^δ and the probability functions which can be obtained from them by the operations of $\oplus_\lambda$ for $\lambda \in [0,1]$ and taking limits.*

4.3 Recoverable Probability Functions

Our aim in this section is to investigate the concept of *Recoverable* probability functions. The idea here is that conditioning on some future observations will lead us back to the probability function we started with, i.e. we can recover this original probability function.

We will show that this concept helps to give an alternative characterisation of the Paris-Nix continuum for $\delta \in [0, 1)$.

Recoverable probability functions are defined as follows.

Definition 4.3.1. We say that a probability function w on L is *Recoverable* if whenever $\phi(a_1, a_2, \dots a_m)$ is a state description that contains the same number of instances of each of the atoms, i.e. if $\phi = \bigwedge_{i=1}^m \alpha_{h_i}(a_i)$ and $m_r = |\{i \mid h_i = r\}| = K$, for $r = 1, \dots 2^p$, $K =$ some constant, then $w(\phi(a_1, a_2, \dots, a_m)) \neq 0$ and

$$w(\theta(a_{m+1}, \dots, a_{m+k}) \mid \phi(a_1, a_2, \dots, a_m)) = w(\theta(a_{m+1}, \dots, a_{m+k})),$$

for all $\theta(a_{m+1}, a_{m+2}, \dots, a_{m+k}) \in SL$.

We will show that it is necessary to have such a restriction on the form of ϕ, that is to say that we cannot just take ϕ to be any state description, but first, we give a rather straightforward lemma that will simplify our later work.

Lemma 4.3.2. *In the definition of Recoverable, it is enough to take $\theta(a_{m+1}, a_{m+2}, \dots, a_{m+k})$ to be a state description. That is to say, If a probability function is Recoverable over all state descriptions, then it is Recoverable over all sentences.*

Proof. Assume we have w such that

$$w(\theta \mid \phi) = w(\theta),$$

where ϕ is some state description which contains the same number of instances of each of the atoms of L, and θ is any state description in L.

Take $\psi \in SL$ to be any sentence, then by the DNFT we can take ψ to be some disjunction of state descriptions. We can therefore express ψ as

$$\psi(a_{m+1}, \ldots, a_{m+k}) = \bigvee_{x=1}^{X} \theta_x(a_{m+1}, \ldots, a_{m+k}),$$

where the θ_x are state descriptions. We want to show that

$$w(\psi \mid \phi) = w(\psi).$$

Using our definition of ψ we re-write this as

$$w\left(\bigvee_{x=1}^{X} \theta_x \mid \phi\right) = w\left(\bigvee_{x=1}^{X} \theta_x\right),$$

which is equivalent to

$$\sum_{x=1}^{X} w(\theta_x \mid \phi) = \sum_{x=1}^{X} w(\theta_x). \tag{4.15}$$

By our initial assumption we have that

$$w(\theta_x \mid \phi) = w(\theta_x)$$

for each choice of x, and therefore we have (4.15). Since we chose ψ arbitrarily the result then follows, as required. □

We now show that in the definition of Recoverable it is necessary to impose the restriction on the form of ϕ if we wish to avoid triviality.

Theorem 4.3.3. *If, in the definition of Recoverable, we required the principle to hold for* **all** *state descriptions, ϕ, then the only probability function which would satisfy Ax and be Recoverable is the Independent Solution, w_I (defined on page 23).*

Proof. Let L contain p predicates, $P_1, \ldots, P_p$, and let $\alpha_1, \ldots, \alpha_{2^p}$ list the atoms of L. Let $\phi(a_1, a_2, \ldots, a_m), \theta(a_{m+1}, a_{m+2}, \ldots, a_{m+k})$ be state descriptions of L.

Throughout (and only in) this proof we take the definition of Recoverable to be the alternative form suggested in the statement of the theorem, which has ϕ as *any* state description.

We first take $w = w_I$ and show that it is Recoverable.

We have, by the definition of w_I, that

$$w_I(\phi(a_1, a_2, \ldots, a_m)) = \frac{1}{2^{mp}},$$

$$w_I(\theta(a_{m+1}, a_{m+2}, \ldots, a_{m+k})) = \frac{1}{2^{kp}},$$

$$w_I(\phi \wedge \theta) = \frac{1}{2^{(m+k)p}},$$

and therefore

$$\begin{aligned} w_I(\theta \mid \phi) &= \frac{w_I(\theta \wedge \phi)}{w_I(\phi)} \\ &= \frac{\frac{1}{2^{(m+k)p}}}{\frac{1}{2^{mp}}} \\ &= \frac{1}{2^{kp}} = w(\theta), \end{aligned}$$

as required.

We now take w to be Recoverable and satisfy Ax.

Recalling our definition of vector representations of state descriptions let

$$\vec{\phi} = \langle n_1, n_2, \ldots, n_{2^p} \rangle,$$

and

$$\vec{\theta} = \langle r_1, r_2, \ldots, r_{2^p} \rangle.$$

Let μ be the de Finetti prior of w, then by Recoverability we have that

$$\int \prod_{i=1}^{2^p} x_i^{n_i + r_i} d\mu = \left(\int \prod_{i=1}^{2^p} x_i^{n_i} d\mu \right) \cdot \left(\int \prod_{i=1}^{2^p} x_i^{r_i} d\mu \right), \tag{4.16}$$

and

$$\int \prod_{i=1}^{2^p} x_i^{r_i} d\mu \neq 0.$$

By repeated use of (4.16) we have that

$$\int \prod_{i=1}^{2^p} x_i^{s_i} d\mu = \prod_{i=1}^{2^p} \left(\int x_i d\mu \right)^{s_i},$$

so by the uniqueness of μ this must mean that μ is discrete and puts all the measure on the single point

$$\left\langle \int x_1\, d\mu,\ \int x_2\, d\mu,\ \ldots,\ \int x_{2^p}\, d\mu \right\rangle.$$

Since w satisfies Ax, each of these must be equal and since they sum to one each must equal $\frac{1}{2^p}$. Our probability function is therefore the one which puts all the measure on the point

$$\left\langle \frac{1}{2^p}, \frac{1}{2^p}, \ldots, \frac{1}{2^p} \right\rangle,$$

i.e. w defines the independent solution, w_I, as required. □

As intimated previously, we can use the Recoverable concept to provide a characterisation of the Paris-Nix continuum, we do so via the main theorem of the section, which now follows.

Theorem 4.3.4. *The probability functions which satisfy Ax, Language Invariance and are Recoverable are precisely the w^δ for $\delta \in [0, 1)$.*

Proof. Let L contain p predicates, $P_1, \ldots, P_p$, and let $\alpha_1, \ldots, \alpha_{2^p}$ list the atoms of L. We now consider each direction in turn.

(1) The w^δ probability functions satisfy Ax, Language Invariance and are Recoverable.

We already know that the Paris-Nix continuum satisfies LIP, and Ax (from [23]), so we only need to show that they are Recoverable[5] for $\delta \in [0, 1)$.

Let $\phi(a_1, a_2, \ldots, a_{2^p})$ be some state description that contains equal numbers of observations of each atom, i.e.

$$\vec{\phi} = \langle K, K, \ldots, K \rangle,$$

where $K = \frac{m}{2^p} \in \mathbb{N}$.

[5] It has since been pointed out that Williamson has now essentially also provided a proof for this direction (only) in [29].

By Lemma 4.3.2 it is enough to take $\theta(a_{m+1},\dots,a_{m+l})$ to be some state description, say

$$\vec{\theta} = \langle n_1, n_2, \dots, n_{2^p}\rangle.$$

Trivially $w^\delta(\theta) \neq 0$, for $\delta \in [0,1)$, so we can concentrate on showing that

$$w(\theta(a_{m+1},\dots,a_{m+l}) \mid \phi(a_1,a_2,\dots,a_m)) = w(\theta(a_{m+1},\dots,a_{m+l})). \tag{4.17}$$

By the definition of w^δ given on page 26 we have that

$$w(\theta) = \frac{1}{2^p}\sum_{i=1}^{2^p}(\gamma+\delta)^{n_i}\gamma^{l-n_i},$$

and

$$\begin{aligned} w(\phi) &= \frac{1}{2^p}\sum_{i=1}^{2^p}(\gamma+\delta)^K\gamma^{m-K} \\ &= \frac{1}{2^p}2^p(\gamma+\delta)^K\gamma^{m-K} \\ &= (\gamma+\delta)^k\gamma^{m-K}. \end{aligned}$$

We have that

$$\begin{aligned} \theta\wedge\phi &= \langle n_1,n_2,\dots,n_{2^p}\rangle + \langle K,K,\dots,K\rangle \\ &= \langle n_1+K, n_2+K,\dots,n_{2^p}+K\rangle, \end{aligned}$$

so

$$\begin{aligned} w(\theta\wedge\phi) &= \frac{1}{2^p}\sum_{i=1}^{2^p}(\gamma+\delta)^{n_i+K}\gamma^{m+l-K-n_i} \\ &= \frac{1}{2^p}(\gamma+\delta)^K\gamma^{m-K}\sum_{i=1}^{2^p}(\gamma+\delta)^{n_i}\gamma^{l-n_i}. \end{aligned}$$

Putting all this together gives

$$\begin{aligned} w(\theta\mid\phi) &= \frac{w(\theta\wedge\phi)}{w(\phi)} \\ &= \frac{\frac{1}{2^p}(\gamma+\delta)^K\gamma^{m-K}\sum_{i=1}^{2^p}(\gamma+\delta)^{n_i}\gamma^{l-n_i}}{w(\phi)} \\ &= \frac{w(\phi)\cdot w(\theta)}{w(\phi)} \\ &= w(\theta), \end{aligned}$$

as required.

(2) We need to prove, then, that if a given probability function, w, is Recoverable, satisfies Ax and LIP, then it is equivalent to w^δ for some $\delta \in [0, 1)$.

Let θ, ϕ be defined as for part (1) of this proof. Let μ be the de Finetti prior of w, then Recoverability gives us

$$\int \prod_{i=1}^{2^p} x_i^{n_i+K} d\mu = \left(\int \prod_{i=1}^{2^p} x_i^{n_i} d\mu \right) \cdot \left(\int \prod_{i=1}^{2^p} x_i^{K} d\mu \right).$$

Suppose that for some $0 < g < 1$ both the sets

$$C_1 = \left\{ \langle x_1, x_2, \ldots, x_{2^p} \rangle \in \mathbb{D}_{2^p} \mid \prod_{i=1}^{2^p} x_i \leq g \right\},$$

$$C_2 = \left\{ \langle x_1, x_2, \ldots, x_{2^p} \rangle \in \mathbb{D}_{2^p} \mid \prod_{i=1}^{2^p} x_i > g \right\},$$

have non-zero μ measure. Let

$$f = \sup \left\{ d \mid \mu \left\{ \langle x_1, x_2, \ldots, x_{2^p} \rangle \in \mathbb{D}_{2^p} \mid \prod_{i=1}^{2^p} x_i \geq d \right\} > 0 \right\},$$

and pick $g < e \leq f$ such that

$$(e - g)\mu(C_1) > (f - e)\mu(C_3)\,, \text{ and } \mu(C_3) > 0, \tag{4.18}$$

where

$$C_3 = \left\{ \langle x_1, x_2, \ldots, x_{2^p} \rangle \in \mathbb{D}_{2^p} \mid e \leq \prod_{i=1}^{2^p} x_i \leq f \right\}.$$

Claim :

$$\lim_{r \to \infty} \frac{\int \prod_{i=1}^{2^p} x_i^r d\mu}{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu} = 1.$$

We take small $\nu > 0$ such that $\mu(C_4) < \nu$ where

$$C_4 = \left\{ \langle x_1, x_2, \ldots, x_{2^p} \rangle \in \mathbb{D}_{2^p} \mid e - \nu \leq \prod_{i=1}^{2^p} x_i < e \right\}.$$

Let C_5 be the complement of $C_3 \cup C_4$, then

$$\begin{aligned}
\frac{\int \prod_{i=1}^{2^p} x_i^r d\mu}{\int_{C_3} \prod_{i=1}^{2^p} \prod_{i=1}^{2^p}} &= \frac{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu \,+\, \int_{C_4} \prod_{i=1}^{2^p} x_i^r d\mu \,+\, \int_{C_5} \prod_{i=1}^{2^p} x_i^r d\mu}{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu} \\
&\leq \frac{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu \,+\, e^r \mu(C_4) + (e-\nu)^r \mu(C_5)}{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu} \\
&\leq 1 + \frac{e^r \mu(C_4)}{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu} + \frac{(e-\nu)^r \mu(C_5)}{\int_{C_3} \prod_{i=1}^{2^p} x_i^r d\mu} \\
&\leq 1 + \frac{e^r \mu(C_4)}{e^r \mu(C_3)} + \frac{(e-\nu)^r \mu(C_5)}{e^r \mu(C_3)} \\
&< 1 + \frac{\nu}{\mu(C_3)} + \frac{(e-\nu)^r \mu(C_5)}{e^r \mu(C_3)} \\
&\leq 1 + \frac{\nu}{\mu(C_3)},
\end{aligned}$$

for large r since $\frac{(e-\nu)}{e^r} \to 0$ as $r \to \infty$ and $\mu(C_4) < \nu$. The result now follows since we can take ν to be arbitrarily small.

Due to this claim, by picking s sufficiently large, we can ensure that

$$(1+\epsilon) \int_{C_3} \prod_{i=1}^{2^p} x_i^s \, d\mu \,\geq\, \int \prod_{i=1}^{2^p} x_i^s \, d\mu,$$

for suitably small ϵ and hence that

$$\frac{\int \prod_{i=1}^{2^p} x_i^{s+1} d\mu}{\int \prod_{i=1}^{2^p} x_i^s d\mu} \geq \frac{\int_{C_3} \prod_{i=1}^{2^p} x_i^{s+1} d\mu}{(1+\epsilon) \int_{C_3} \prod_{i=1}^{2^p} x_i^s d\mu}. \tag{4.19}$$

Since $\prod_{i=1}^{2^p} x_i \geq e$ on C_3 we have that

$$\frac{\int_{C_3} \prod_{i=1}^{2^p} x_i^{s+1} d\mu}{(1+\epsilon) \int_{C_3} \prod_{i=1}^{2^p} x_i^s d\mu} \geq \frac{\int_{C_3} e \prod_{i=1}^{2^p} x_i^s d\mu}{(1+\epsilon) \int_{C_3} \prod_{i=1}^{2^p} x_i^s d\mu} = \frac{e}{1+\epsilon}, \tag{4.20}$$

and note that

$$\frac{e}{1+\epsilon} = \frac{e\mu(C_1) + e\mu(C_2 - C_3) + e\mu(C_3)}{1+\epsilon}, \tag{4.21}$$

we know that

$$\int \prod_{i=1}^{2^p} x_i \, d\mu \,\leq\, g\mu(C_1) + e\mu(C_2 - C_3) + f\mu(C_3), \tag{4.22}$$

because $\prod_{i=1}^{2^p} x_i \leq g, e, f$ on $C_1, (C_2 - C_3)$ and C_3 respectively. Noting that

$$\frac{\int \prod_{i=1}^{2^p} x_i^{s+1} d\mu}{\int \prod_{i=1}^{2^p} x_i^s d\mu} = \int \prod_{i=1}^{2^p} x_i \, d\mu,$$

by Recoverability, and putting (4.19), (4.20), (4.21) and (4.22) together we get that

$$\frac{e\mu(C_1) + e\mu(C_2 - C_3) + e\mu(C_3)}{1+\epsilon} \leq g\mu(C_1) + e\mu(C_2 - C_3) + f\mu(C_3), \qquad (4.23)$$

we can take ϵ to be arbitrarily small, so (4.23) must hold for $\epsilon = 0$. Therefore we have that this reduces to

$$\begin{aligned} e\mu(C_1) - g\mu(C_1) &\leq f\mu(C_3) - e\mu(C_3) \\ \Rightarrow (e-g)\mu(C_1) &\leq (f-e)\mu(C_3), \end{aligned}$$

contradicting (4.18). Therefore there can be no such g and, in fact, there must be some $0 \leq d \leq 1$ such that

$$\mu\left\{\langle x_1, x_2, \ldots, x_{2^p}\rangle \in \mathbb{D}_{2^p} \mid \prod_{i=1}^{2^p} x_i = d\right\} = 1. \qquad (4.24)$$

We also have that $d \neq 0$, by the following argument.

Let $\phi(a_1, a_2, \ldots, a_m)$ be a state description of L that contains an equal number of observations, k say, of each atom then, by Recoverability, $w(\phi) \neq 0$. However, if all the measure was placed on points whose product of coordinates was 0 (i.e. $d = 0$) we'd have that $w(\phi) = 0$, thus forming a contradiction. Therefore $d \neq 0$.

We next consider this in relation to Language Invariance. We take our primary language as defined at the beginning of the proof, with L^+ defined to be $L \cup \{P_{p+1}(x)\}$. Let $\alpha_1^1, \alpha_1^0, \alpha_2^1, \ldots, \alpha_i^1, \alpha_i^0, \ldots, \alpha_{2^p}^0$ list the atoms where

$$\alpha_i^\epsilon = \alpha_i \wedge P_{p+1}^\epsilon.$$

Let w^+ satisfy Ax and extend w on L^+ and let μ^+ be the corresponding de Finetti prior. Then by (4.24) we have that

$$\mu^+\left\{\vec{x} \in \mathbb{D}_{2^{p+1}} \mid (x_1 + x_2)(x_3 + x_4)\prod_{i=3}^{2^p}(x_{2i-1} + x_{2i}) = d\right\} = 1,$$

for some constant $0 < d \leq 1$, and by Ax

$$\mu^+\left\{\vec{x} \in \mathbb{D}_{2^{p+1}} \mid (x_1 + x_3)(x_2 + x_4)\prod_{i=3}^{2^p}(x_{2i-1} + x_{2i}) = d\right\} = 1,$$

$$\mu^+ \left\{ \vec{x} \in \mathbb{D}_{2^{p+1}} \mid (x_1 + x_4)(x_2 + x_3) \prod_{i=3}^{2^p} (x_{2i-1} + x_{2i}) = d \right\} = 1.$$

Hence

$$\mu^+ \{ \vec{x} \in \mathbb{D}_{2^{p+1}} \mid (x_1 + x_2)(x_3 + x_4) = (x_1 + x_3)(x_2 + x_4) = (x_1 + x_4)(x_2 + x_3) \} = 1.$$

We next note that

$$\begin{aligned}
(x_1 + x_2)(x_3 + x_4) &= (x_1 + x_3)(x_2 + x_4) \\
&\iff \\
x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 - x_1x_2 - x_1x_4 - x_3x_2 - x_3x_4 &= 0 \\
\iff x_1x_3 + x_2x_4 - x_1x_2 - x_3x_4 &= 0 \\
\iff (x_1 - x_4)(x_3 - x_2) &= 0,
\end{aligned}$$

and similarly,

$$(x_1 - x_3)(x_4 - x_2) = (x_1 - x_2)(x_4 - x_3) = 0,$$

therefore

$$\mu^+ \{ \vec{x} \in \mathbb{D}_{2^{p+1}} \mid (x_1 - x_4)(x_3 - x_2) = (x_1 - x_3)(x_4 - x_2) = (x_1 - x_2)(x_4 - x_3) = 0 \} = 1. \tag{4.25}$$

Claim :

We claim that the x_i's can only take one of two possible values, and that one of those values can only be taken at most once.

Let the x_i's take more than two possible values, then it is possible to find x_j, x_k, x_r such that

$$x_j \neq x_k \neq x_r.$$

By Ax we can take $x_1 = x_j$, $x_2 = x_k$, $x_3 = x_r$. We have that

$$(x_1 - x_4)(x_2 - x_3) = 0,$$

from (4.25), so $x_1 - x_4$ must equal 0, and therefore $x_1 = x_4 \neq x_3 \neq x_2$, but then

$$(x_1 - x_3)(x_4 - x_2) \neq 0,$$

contradiction, therefore the x_i's can take at most two different values.

If the x_i's could take both of these values more than once, then we'd be able to find x_j, x_k, x_r, x_s such that

$$x_j = x_k,\ x_r = x_s,\ x_j \neq x_r.$$

By Ax we can take $x_1 = x_j$, $x_2 = x_k$, $x_3 = x_r$, $x_4 = x_s$. Then

$$(x_1 - x_3)(x_4 - x_2) \neq 0,$$

contradiction, therefore one of the two values can be taken at most once, as required.

Just to finish up, we show that the conditions in (4.25) are not violated by taking all x_i's equal except for one. By Ax we can take x_1 to be the different coordinate, with $x_2 = x_3 = x_4$ then we have that

$$(x_1 - x_4)(x_3 - x_2) = (x_1 - x_3)(x_4 - x_2) = (x_1 - x_2)(x_4 - x_3) = 0,$$

as required. Note that if all the values are the same, the condition is trivially satisfied, so we did need to say "at most once" in the claim.

Our claim states that all the measure is put upon vectors in the set of $\vec{x}$ for which all except (possibly) one of the x_i are equal. Let this x_i be represented by x_I. Say that the rest of the x_i's are equal to $x \in [0, \frac{1}{2^{p+1}-1}]$ then

$$x_I = (1 - (2^{p+1} - 1)x).$$

We know that for some $0 < d \leq 1$ we have that

$$d = (2x)^{2^p - 1}(1 - (2^{p+1} - 2)x),$$

meaning that, up to permutation of coordinates, there are two such possible $\vec{x}$, one where $x_I > x$ and the other where $x_I < x$ (if $x_I = x$ the two possibilities coincide). These possibilities are of the form

$$\langle x, x, \ldots, x, x_I, x, \ldots x, x\rangle.$$

Notice that, by Ax, all the points in each of these two families receive the same μ^+ measure.

Our aim now is to show that the case $x_I < x$ is not possible (i.e. each such point receives 0 measure). Suppose otherwise, then by the same argument as above we could show that the corresponding μ^{++} on $L \cup \{P_{p+1}(x) \wedge P_{p+2}(x)\}$ must give non-zero measure to points

$$\langle y, y, \ldots, y, y_I, y, \ldots, y, y \rangle,$$

where $y + y = x$, $y + y_I = y_I$, then

$$y_I = x_I - \frac{x}{2}.$$

Similarly for μ^{+++} we have (with z_I as appropriate)

$$z_I = x_I - \frac{x}{2} - \frac{x}{4},$$

and if we keep extending our language we'll end up with our 'different' coordinate, σ say, to be

$$\sigma = x_I - \left(\sum_{r=1}^{R} \frac{x}{2^r} \right) \Rightarrow_{R \to \infty} x_I - x < 0,$$

which is impossible, therefore a contradiction is formed and we can concentrate solely on the case where $x_I \geq x$.

So we now know that all the measure is spread uniformly over the 2^{p+1} permutations of the point

$$\langle x, x, \ldots, x, x_I, x, \ldots, x, x \rangle,$$

where $x_I > x$, hence by marginalising the same is true in our original language L, that is to say all μ's measure is uniformly distributed over the point

$$\langle \gamma, \gamma, \ldots, \gamma, \tau, \gamma, \ldots, \gamma, \gamma \rangle, \tag{4.26}$$

for some $\tau \geq \gamma$. Hence setting $\tau = \gamma + \delta$ proves our claim.

We now have that $w = w^\delta$ for some $\delta \in [0, 1]$, we complete our proof by noting that 23 must have that $w(\vec{k}) \neq 0$ (since w is Recoverable). This specifically rules out any vector where any of the coordinates are 0. Trivially this can only happen in 4.26 when $\delta = 1$, and therefore $w = w^\delta$ for some $\delta \in [0, 1)$, as required. $\square$

Chapter 5

Conclusions & Suggestions For Future Work

In this chapter we sum up the results found throughout the thesis and suggest some possible avenues for future research.

5.1 The Only Rule

In our introduction we posed the question (in relation to the urn model, see page 10);

"Should we assign greater belief to picking 4 red and 1 green to picking 3 red and 2 green (in that order) in our first five picks?"

The answer to this question (which is 'Yes'), and others like it, is given by the Only Rule. We also showed that if we had asked

"Should we assign greater belief to picking 5 red, 3 blue, 3 green and 1 yellow or to 4 black, 4 yellow, 4 red and no orange (in that order) in our first ten picks?"

Then no definitive answer can be given (where we assume Ax).

This is of interest as it identifies some of the underlying structure that we assume when we take our probability functions to satisfy Ax. It is interesting to note that these probability functions are biased in favour of expecting mono-observational systems, though we might expect this given that the whole point of inductive learning is to base future expectations on what you have already seen!

The usefulness of this result is immediately apparent, in its use in some of the proofs in the following chapter on Relevance and, it is hoped, in forthcoming work.

The "Long Road" proof of The Only Rule is very illuminating since it demonstrates that the result holds essentially because when $\theta = \langle n_1, \ldots, n_{2^p} \rangle$ and $\phi = \langle m_1, \ldots, m_{2^p} \rangle$ are state descriptions such that

$$\sum_{j \geq i} m_j \geq \sum_{j \geq i} n_j \quad \text{for } i = 1, 2, 3 \ldots 2^p$$

then for some large K there are more non-trivial[1] extensions of θ to a state description containing K observations, than there are of ϕ. This fact is hidden within the more succinct, neater "Short-Cut".

In terms of suggestions for future work, we feel that an n-ary version of the Only Rule would be of benefit - indeed we believe some (unpublished) work has already been done in this area. The benefits of the "Long Road" proof again rears its head here, since there does not exist[2] an n-ary version of de Finetti's theorem, so it would appear that the Long Road can more readily be generalised.

Other future work in this area could be done by looking at what happens with regards other types of relationship, for instance, what can be said about the inequality

$$\begin{aligned} w\left(\langle n_1+2, n_2, \ldots, n_{2^p} \rangle\right) - w\left(\langle n_1+1, n_2, \ldots, n_{2^p}+1 \rangle\right) \\ \geq w\left(\langle n_1+1, n_2, \ldots, n_{2^p}+1 \rangle\right) - w\left(\langle n_1, n_2, \ldots, n_{2^p}+2 \rangle\right), \end{aligned}$$

and other relationships of its ilk?

[1]Non-trivial is here intended to mean that the probability of the extension does not tend to zero as we increase K.

[2]See the footnote on page 34.

5.2 Principles Involving Relevance

5.2.1 NPIR & SPIR

In our introduction we posed the question;

"After we have picked 2 reds and a green should we give greater expectation to the 5th ball being red if the 4th ball is red rather than green?"

and the surprising answer given by the work in this chapter is that, when we assume just Ax, "Not Always". A similar question which has the same answer is

"After we have picked 2 reds and a green should we give lower expectation to the 5th ball being green once we have made a further observation of a red ball."

This latter example relates to the principle NPIR.

This is not what we would expect from ostensibly sensible principles (though the failures occur with what would appear to be undesirable probability functions). This is summed up and discussed in the discussion section (pp.90-91), so we will not repeat ourselves here.

In terms of possible future work the obvious first thing would be to try and close the unresolved cases so that we have a complete overview of where these principles hold and where they do not. One, seemingly hard, problem which may entice future work is whether a continuum of probability functions can be found for these principles (such as with the Paris-Nix Continuum for GPIR). It may also be worth looking into and adopting other relevance principles, though this immediately gives us the philosophical problem of choosing which one(s)! Finally, it will be of interest to follow up on what happens when we apply these principles to n-ary languages.

5.2.2 PIR & Ex

It has long been known that $Ex \rightarrow PIR$, and here we showed that the reverse is not true, albeit for a version of PIR which is not the most general.

The result here is of merit however, since the result holds for the version of PIR which most closely matches the intuition behind the principle - namely that once we have observed something, we should give a greater expectation to observing it again.

The most general version, however, seems to be a lot woolier in its verbal form. Namely, returning to our urn example, we have that

> "If we pick out a red, yellow or orange ball then we should assign greater probability to the next ball being red, yellow or orange",

but the initial idea of PIR is that we already *know* what our additional observation is.

However, it may well be of interest to discover what the situation is regarding this most general form of PIR and Ex, so we suggest this as an idea for future research.

5.3 Principles Involving Irrelevance

5.3.1 The Weak Irrelevance Principle

In our introduction we posed the question (upon splitting our coloured balls amongst two different urns)

> "Should the expectation of the next ball drawn from the first urn being red be unchanged, whatever we've drawn from the second urn?"

the (perhaps surprising) answer, which was already known prior to the work in this thesis, is usually "No".

Our motivation was to try and discover which probability functions, also satisfying Ax, do satisfy this principle. Whilst we do not answer this fully, we do provide a conjecture that the family of probability functions we generate in that chapter are the only probability functions which satisfy WIP, LIP, Ex and Ax.

Our suggestion for future work here then is clear, to prove the conjecture we offer.

5.3.2 Recoverable Probability Functions

In terms of our urn model, an example of this principle would be

> "We should give the same probability to the next ball being red, whether we have observed nothing at all, or if we have observed one of each of the colours."

We show that it is necessary to state that we have observed equal numbers of each of the balls (i.e. that we cannot replace the condition with an arbitrary set of observations), otherwise the only probability function which satisfies Ax and is Recoverable is the trivial Independent Solution, w_I.

The main result of this chapter is then given. We show that the Recoverable functions, that also satisfy Ax and LIP, are exactly those given by the Paris-Nix Continuum, thus giving an alternative characterisation of these functions which had been previously characterised by GPIR.

Our only suggestion for future work here, is to see how this principle operates in languages with n-ary predicates, whether the same result holds for extensions of GPIR and the Paris-Nix Continuum (if these exist!).

5.4 Overall

The aims for our thesis, as stated in section 1.4 were, in brief,

(1) To uncover some of the underlying structure behind symmetric probability functions.
(2) To give insight into principles which have not been studied in depth before.
(3) To answer some unanswered questions within the field.

(4) To provide a platform for future work.

We feel that all of these aims have been met, the first through the work in Chapter 2 on The Only Rule. The second through our work in Chapters 3 and 4 on SPIR, NPIR, WIP and Recoverable probability functions. The third through all of the thesis, but most notably the work on PIR and Ex in section 3.5. Finally the fourth aim throughout the work, as shown by the suggestions made above.

Overall then we do feel that we have addressed several facets of the area of Uncertain Reasoning, pushing each forward a little way in order to provide a significant overall contribution.

Bibliography

[1] R. Audi (ed.) - The Cambridge Dictionary of Philosophy *Cambridge University Press, 1995.*

[2] R. Carnap - On the Application of Inductive Logic *Philosophy and Phenomenology Research, 1947, 8:133-147.*

[3] R. Carnap - Logical Foundations of Probability *Routledge & Kegan Paul Ltd. 1951.*

[4] R. Carnap - The Continuum of Inductive Methods *University of Chicago Press, 1952.*

[5] R. Carnap - Replies and Systematic Expositions *The Philosophy of Rudolf Carnap, ed. P.A.Schipp, La Salle, Illinois, Open Court, 1963, pp.859-999.*

[6] R. Carnap - A Basic System of Inductive Logic, Part I *Studies in Inductive Logic and Proability, eds. R.Carnap, R.Jeffrey, University of California Press, 1971, pp.36-165.*

[7] W. Feller - An Introduction to Probability Theory and its Applications, Volume II, Second Edition, *John Wiley and sons Inc. 1971.*

[8] B. de Finetti - Sul significato soggetivo della probabilità *Fundamenta Mathematicae, 1931, Vol.17, pp.298-329.*

[9] B. de Finetti - La Prévision: Ses Lois Logiques, Ses Sources Subjectives, *Annales de l'Institut Henri Poincaré, 7, 1-68, 1937.*

[10] B. de Finetti - Theory of Probability, Vol 1, *Wiley, New York, 1974.*

[11] B. de Finetti - On the Condition of Partial Exchangeability, *Studies in Inductive Logic and Proabbility, Volume II, ed. R.C. Jeffrey, University of California Press, Berkley and Los Angeles, 1980.*

[12] H. Gaifman - Concerning Measures in First Order Calculi, *Israel Journal of Mathematics, 1964, vol.2, pp 1-18.*

[13] H. Gaifman - Applications of de Finetti's Theorem to Inductive Logic *Studies in Inductive Logic and Probability, eds. R.Carnap, R.Jeffrey, University of California Press, 1971, pp.235-251.*

[14] S.Glaister - Inductive Logic, *A Companion to Philosophical Logic, ed. D.Jacquette, Blackwell, London, 2001, pp.565-581.*

[15] N. Goodman - Fact, Fiction and Forecast, *London : Athlone Press, 1954.*

[16] J. Hintikka and I. Niiniluoto *An Axiomatic Foundation for the Logic of Inductive Generalization, in R.C. Jeffrey (ed.), Studies in Inductive Logic and Probability, Volume 2, University of California Press, Berkeley, CA, 1980, pp. 157-181.*

[17] M. J. Hill - Aspects of Induction and the Principle of Duality. *PhD Thesis, University of Manchester, 1999*

[18] M. J. Hill, J. B. Paris and G. Wilmers - Some observations on Induction in Predicate Probabilistic Reasoning. *Journal of Philosophical Logic, 31:43-75, February 2002*

[19] J. Hümburg - The Principle of Instantial Relevance *Studies in Inductive Logic and Proability, eds. R.Carnap, R.Jeffrey, University of California Press, 1971, pp.225-234.*

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

[21] K. Kelly - The Logic of Reliable Inquiry *Cambridge, Cambridge University Press, 1996.*

[22] E.Mendelson - Introduction to Mathematical Logic *Wadsworth Inc, 1987.*

[23] C.J. Nix - Probabilistic Induction in the Predicate Calculus *PhD Thesis, University of Manchester, 2005*, http://www.maths.manchester.ac.uk/~jeff/theses/chris-nix.ps.

[24] C.J. Nix and J.B. Paris - A Note on Binary Inductive Logic *to appear in the Journal of Philosophical Logic.*

[25] C.J. Nix and J.B. Paris - A Continuum of Inductive Methods Arising From a Generalized Principle of Instantial Relevance, *Journal of Philosophical Logic, Online First Issue, DOI:10,1007/s 10992-005-9003x, ISSN 0022-3611 (Paper) 1573-0433, 2005.*

[26] J.B. Paris - The Uncertain Reasoners Companion *Cambridge University Press, 1994.*

[27] J.B. Paris - Non-monotonic Logic Course Notes http://www.maths.manchester.ac.uk/~jeff/lecture-notes/mt4181.ps.

[28] F.P. Ramsey - The Foundations of Mathematics and Other Logical Essays, *London, 1931.*

[29] J. Williamson - Inductive Influence, *to appear in the British Journal for the Philosophy of Science.*

[30] S. Zabell - Confirming Universal Generalisations *Erkenntnis, Volume 45, Numbers 2-3, pp. 267-283, 1996.*

[31] S. Zabell - The Continuum of Inductive Methods Revisited *J. Earman and J. Norton (eds), The Cosmos of Science (pp. 351-385), Pittsburgh, University of Pittsburgh Press, 1997.*

[32] S. Zabell - Symmetry and Its Discontents : Essays on the History of Inductive Probability *Cambridge University Press, 2005.*

Appendix A

Mathematica Programme for Checking SPIR Counter-Example

In Chapter 3 we used the following Mathematica programme to check our Counter-Example to SPIR for the case where $n_3, n_4, \ldots, n_{2^p}$ are not all 0.

```
V = 0.55;
T = 0.44; (* tau *)
F = (1 - V - T) / 6; (* epsilon, e is protected in mathematica *)
Array[n, 8]; (* n vector *)
n[1] = 17; n[2] = 21; n[3] = 3; n[4] = 2; n[5] = 2; n[6] = 2;
n[7] = 1; n[8] = 1;
(* These are slightly different to how they appear in the
   notes in order to get the i ≠ j part involved *)
```

$$x1 = 6!\sum_{i=2}^{8}\left(T^{n[i]}\left(\frac{\prod_{j=2}^{8} F^{n[j]}}{F^{n[i]}}\right)\right);$$

$$x2 = 6!\sum_{i=2}^{8}\left(V^{n[i]}\left(\frac{\prod_{j=2}^{8} F^{n[j]}}{F^{n[i]}}\right)\right);$$

```
x3 = 0 ; (* need a for loop to do this one as we need k ≠ i,j *)
For [i = 2, i < 9, i++,
 For[j = 2, j < 9, j++,
  If[i ≠ j,
   {NT = 0;
    For[k = 2, k < 9, k++,
     If[k ≠ i && k ≠ j,
      NT = NT + n[k]]];
    x3 = x3 + V^n[i] T^n[j] F^NT}, 0]]];
x3 = 5 ! x3;
n[2] = n[2] + 1;
(* y1, y2, y3 are x' _1 x' _2, x' _3 in the notes *)
```

$$y1 = 6\,!\sum_{i=2}^{8}\left(T^{n[i]}\left(\frac{\prod_{j=2}^{8}F^{n[j]}}{F^{n[i]}}\right)\right);$$

$$y2 = 6\,!\sum_{i=2}^{8}\left(V^{n[i]}\left(\frac{\prod_{j=2}^{8}F^{n[j]}}{F^{n[i]}}\right)\right);$$

```
y3 = 0 ; (* need a for loop to do this one *)
For [i = 2, i < 9, i++,
 For[j = 2, j < 9, j++,
  If[i ≠ j,
   {NT = 0;
    For[k = 2, k < 9, k++,
     If[k ≠ i && k ≠ j,
      NT = NT + n[k]]];
    y3 = y3 + V^n[i] T^n[j] F^NT}, 0]]];
y3 = 5 ! y3;

Print["X 1 = "]
XX1 = V^(n[1]+2) x1 + T^(n[1]+2) x2 + 6 F^(n[1]+2) x3
Print["X 2 = "]
XX2 = V^(n[1]+1) x1 + T^(n[1]+1) x2 + 6 F^(n[1]+1) x3
Print["X 3 = "]
XX3 = V^(n[1]+1) y1 + T^(n[1]+1) y2 + 6 F^(n[1]+1) y3
Print["X 4 = "]
XX4 = V^n[1] y1 + T^n[1] y2 + 6 F^n[1] y3
Print["Left Hand Side ="]
LHS = 8 * XX1 + XX4
Print["Right Hand Side ="]
RHS = 8 * XX3 + XX2
Print["Left Hand - Right Hand = "]
Simplify[RES = LHS - RHS]
```