

A Note on Binary Inductive Logic

Nix, C. J. and Paris, J. B.

2008

MIMS EPrint: **2008.106**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

A Note on Binary Inductive Logic

C. J. Nix* and J. B. Paris

School of Mathematics

University of Manchester

Manchester M13 9PL

chris.nix@gmail.com jeff.paris@manchester.ac.uk

August 15, 2008

Abstract

We consider the problem of induction over languages containing binary relations and outline a way of interpreting and constructing a class of probability functions on the sentences of such a language. Some principles of inductive reasoning satisfied by these probability functions are discussed, leading in turn to a representation theorem for a more general class of probability functions satisfying these principles.

Introduction

Despite the expectation expressed by Kemeny in his paper [23] written in 1954 *Inductive Logic*, based as in the early work of Johnson [19], Carnap [1], [2] et al on symmetry principles, appears up to this time to have been, with the exception of Krauss's [24] and early work of Hoover [17], almost wholly concerned with unary languages and induction for *unary* predicates (see, for example, the summary paper [12]¹). However in everyday life inductive reasoning is certainly not entirely limited to just unary languages. We on occasions seem to make inductions from knowledge in which binary relations, or perhaps even relations of greater arity, are present. For example a gardener, Adam say, has noticed that certain apple trees have the ability to pollinate other apple trees, even to self pollinate. It

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

¹Apparently largely unbeknown to the Inductive Logic community there has however been a substantial development beyond the unary in the area of Probability Theory, in particular in proving de Finetti style theorems for what are known there as 'jointly exchangeable arrays'. We refer the reader to [20], [21].

does not take much to suppose that if Adam has noticed that apple variety A is generally a good pollinator and variety B easily pollinated that he might recommend planting them together even though he had no prior experience of the efficacy of their cultivation in direct proximity.

Our plan in this note is to make some initial observations on the ‘problem of induction’, i.e. what measure of belief to assign to unknown events on the basis of past experience, in the case of purely binary predicate languages.

We shall start by considering some properties we might expect of such a measure. We shall then suggest a *putative* explanation of what it means for a binary relation to hold between two individuals and show that a consequence of this assumption, Spectrum Exchangeability, implies these expected properties as well as being an attractive, and apparently widely generalizable, property in its own right. We shall then go on to derive a characterization of a subclass of probability functions satisfying Spectrum Exchangeability.

Throughout we shall work with languages L for predicate logic with (just) the constant symbols a_1, a_2, a_3, \dots and *finitely* many predicate symbols, but without function symbols or equality. Our intended interpretation here is that these constants exhaust the universe. Let FL denote the formulae of L , SL the sentences of L (i.e. closed formulae) and $QFFL/QFSL$ the quantifier free formulae/sentences of L .

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

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

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

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

Throughout w, v , possibly with various annotations, will denote probability functions (on the ‘obvious’ languages) and, for the purposes of motivation, we shall be thinking of probabilities in the sense of de Finetti as subjective degrees of belief².

Given a probability function w on L we define, as usual, a corresponding two place *conditional probability function* on $SL \times SL$, denoted by $w(\cdot|\cdot)$, to be such that

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

²Despite *motivating* the principles we introduce by appealing to (hopefully) shared intuitions we would not wish to claim anything more of this paper than that it gives a number of mathematical results of the form ‘if you accept *this* then *that* is a mathematical consequence’. In particular we will not get into the mine field of specifying further what sort of probability, rationality, truth we are supposed to be dealing with. It is up to the reader to decide whether the ‘this’ is relevant to their personal angle on the so called problem of induction.

whenever $\theta, \phi \in SL$ and $w(\phi) > 0$. We shall take $w(\theta|\phi)$ to be unspecified when $w(\phi) = 0$. All the elementary properties and conventions concerning probability and conditional probability functions will be assumed (see for example [32]).

One somewhat less elementary result which it will be useful to note is the theorem of Gaifman (see [10], where in fact the axioms (P1-3) were first formulated) that any probability function defined on $QFSL$ (i.e. satisfying (P1) and (P2) for $\theta, \phi \in QFSL$) extends uniquely to a probability function on L . In this sense then we can largely limit our considerations to probability functions defined just on $QFSL$. Furthermore, in that case $w(\theta)$ is determined by its values on the *state descriptions*, that is sentences of the form

$$\bigwedge_j \bigwedge_{b_1, \dots, b_{r_j} \in \{a_{t_1}, \dots, a_{t_m}\}} \pm Q_j(b_1, \dots, b_{r_j}),$$

where the r_j -ary predicates Q_j and the constants a_{t_1}, \dots, a_{t_m} include all those predicates and constants appearing in θ . (For more details within our current notation see [32], and for extensions to richer languages see [11], [34]).

In particular if L is a unary language, that is all the predicates appearing in L are unary, then w is determined by its values on the sentences of the form

$$\bigwedge_{i=1}^p \alpha_{h_i}(a_i)$$

where $1 \leq p \in \mathbb{N}$ and the $\alpha_n(x)$ run through the *atoms* with respect to the set P_1, \dots, P_n of unary predicates from L , that is the 2^n formulae of the form

$$\bigwedge_{j=1}^n P_j^{\epsilon_j}(x)$$

where the $\epsilon_j \in \{0, 1\}$ and $P^1 = P$, $P^0 = \neg P$.

Similarly if L is a binary language then w is determined by its values on the sentences

$$\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j)$$

where the $\beta_r(x, y)$ run through the *atoms* with respect to the set R_1, \dots, R_m of binary predicates from L , that is the 2^m formulae of the form

$$\bigwedge_{j=1}^m R_j^{\epsilon_j}(x, y).$$

In what follows n and m will, as far as possible, be fixed in their roles as the number of unary/binary predicates in the unary/binary language under consideration.

Within this framework the basic ‘problem of induction’ is what belief should be given to a sentence θ *solely* on the basis of knowledge ϕ ? Here ‘should’ is intended as an appeal for justification rather than implying that there is some unique correct answer. In other words, interpreting this according to the ‘received view’ (see [9]), we wish to somehow justify a value for $w(\theta|\phi)$ for w a probability function corresponding to some agent’s beliefs. However since this is determined by $w(\theta \wedge \phi)$ and $w(\phi)$ (assuming $w(\phi) \neq 0$ ³) this question really reduces to justifying a particular choice of a probability function w on L *in the absence of any further knowledge*. As we have already noted this in turn reduces to the problem of justifying probability values for state descriptions alone, a fact that we shall assume without further mention in what follows.

It is crucially important here to emphasize the condition *in the absence of any further knowledge* because the persistent failure within the area of uncertain reasoning to grasp this point has resulted in much misunderstanding and missed opportunity. Examples such as Goodman’s Grue Paradox, [13] (or [35] for a survey) which rely on considerable additional background knowledge are irrelevant as far as the above question is concerned⁴ except in as far as one feels that the same consequence would be justified even without the additional background information. This clearly is not the case in the Grue Paradox where one’s prior knowledge that the color in emeralds does not change over time but that the calendar year we are in does is germane to the ‘induction’, see for example [22]. Indeed without this prior knowledge there would simply be no justification for treating green and grue in any way differently.

Of course one may object that as far as ‘applied philosophy’ is concerned this requirement of total evidence is a very limiting one which in practice is almost never fulfilled⁵, and by the same score should make our own real world, motivating, examples, such as Adam’s apples, worthless as far as we are concerned! Nevertheless with some practice we do seem to be able to imagine, share intuitions about, and on occasions see real world approximations to, such a state of pure ignorance, a state where logic alone is all we have to generate probabilities

³For a detailed consideration of the case when $w(\phi) = 0$ see [6].

⁴In §45B of [1] Carnap is at pains to point out this absence of further knowledge requirement, or as he calls it *requirement of total evidence*, and again in [4] as his *Principle of Total Evidence*. Notwithstanding his appeal to this principle in his defence in [4] to the Grue Paradox (see [14], also [15], [22]) he subsequently seems to abandon it in [5], instead stepping into the semantic minefield of classifying the ‘inductive projectibility’, or otherwise, of real world unary predicates. One can only assume that Carnap came to side with Goodman’s view in [15] (see [18]) that the requirement of total evidence in its widest sense was just too demanding a condition to put on a system which he intended to be generally and practically *applicable* in the real world. Whatever his reasons the Grue Paradox subsequently had a major influence on Carnap and philosophers in general and seem to have been an important contributing factor in the widespread abandonment of the program we are advocating here of studying induction via normative principles in a highly formal and idealized setting.

⁵Except perhaps in the world of expert systems and artificial agents.

on a formal language devoid of any intended interpretations. This very fact suggests to us that it is well worth exploring this situation, just as in Mathematics one might study the Theory of Infinite Sets, and this is the viewpoint we shall adopt in this paper.

Returning to the above ‘problem of induction’ there are certain further properties that it would seem very reasonable to demand of such a probability function given that the assumed state of ignorance gives no cause to treat any of the constants differently from any other, nor any two predicate symbols of the same arity, nor even a predicate and its negation. This leads to us limiting ourselves throughout to probability functions w on languages L satisfying:

The Constant Exchangeability Principle (Ex)⁶

For $\theta, \theta' \in QFSL$, if θ' is obtained from θ by replacing, respectively, the (distinct) constant symbols $a_{i_1}, a_{i_2}, \dots, a_{i_m}$ occurring in θ by the (distinct) constant symbols $a_{k_1}, a_{k_2}, \dots, a_{k_m}$ then $w(\theta) = w(\theta')$.

The Predicate Exchangeability Principle (Px)

For $\theta, \theta' \in QFSL$, if θ' is obtained from θ by replacing, respectively, the (distinct) predicate symbols $Q_{j_1}, Q_{j_2}, \dots, Q_{j_m}$ occurring in θ by the (distinct) predicate symbols $Q_{s_1}, Q_{s_2}, \dots, Q_{s_m}$ of equal arity then $w(\theta) = w(\theta')$.

Strong Negation Principle (SN)

For $\theta, \theta' \in SL$, if Q is any predicate symbol of L and θ' is obtained from θ by replacing each occurrence of Q in θ by $\neg Q$ then $w(\theta) = w(\theta')$.

While we are listing principles it will be useful to also mention:

Regularity Principle (REG)

For consistent $\theta \in QFSL$, $w(\theta) > 0$.

Thus if w satisfies REG it never dismisses a quantifier free sentence θ as impossible (i.e. $w(\theta) = 0$) unless it is forced to (by (P1),(P2)) because θ is inconsistent.

⁶Contra to this general neglect, in the case of a language with single binary predicate this principle was studied by Krauss in [24] where he proved in particular an interesting analogue of de Finetti’s Representation Theorem, see [7], [8].

Spectrum Exchangeability

From now on assume that we are working with a purely binary language⁷, so, recalling the earlier notation, the atoms with respect to the (binary) predicates R_1, \dots, R_m of the language are the formulae

$$\beta_r(x, y) = \bigwedge_{j=1}^m R_j^{\epsilon_j}(x, y),$$

for $r = 1, \dots, 2^m$. We will on occasions refer to a probability function on such a binary language as a *binary probability function* (and similarly for unary).

In the case of such probability functions there are two further principles that would seem justified, as with Ex, Px and SN⁸, on grounds of there being ‘no reason to distinguish’.

The Variable Exchangeability Principle (Vx)

For $\theta, \theta' \in QFSL$ and $1 \leq k \leq m$ if θ' is obtained from θ by replacing each occurrence of $R_k(a_i, a_j)$ in θ by $R_k(a_j, a_i)$ for each i, j then $w(\theta) = w(\theta')$.

The Principle of Unary Conformity (UC)

For $\vec{r} \in \{1, 2, \dots, 2^m\}^p$ and fixed k ,

$$w\left(\bigwedge_{i=1}^p \beta_{r_i}(a_i, a_i)\right) = w\left(\bigwedge_{i=1}^p \beta_{r_i}(a_i, a_k)\right) = w\left(\bigwedge_{i=1}^p \beta_{r_i}(a_k, a_i)\right)$$

Note that the second equality in UC actually follows from Vx. To motivate the principle UC consider the sentences formed just using $R_1(x, x)$. In this case we would effectively be back in the situation of having a unary language. Indeed it would be hard to argue that whatever the choice of w its restriction to this language should not be the same as the choice which would have been adopted on a unary language in the first place. But now for a fixed a_k the same (hand waving!) remarks would seem equally to apply to the $R_1(x, a_k)$, and similarly to the $R_1(a_k, x)$. In both cases it might be argued that the restriction of w to the sentences formed from these *unary* predicates should match the choice would have been made on a unary language. Of course we still seem a long way from saying what this choice should be, but nevertheless we can still draw the conclusion, as expressed by UC, that these three restrictions of w *should all be the same*.

⁷All of the results we present here have corresponding versions in the case of a general language with finitely many, not necessarily binary, relations, for details see the forthcoming [25]. However the notation and visualization required becomes formidable without, apparently, yielding any significantly greater insights. Largely for that reason we restrict ourselves in this paper to a purely binary language.

⁸It can be shown that these principles do not follow from Ex, Px and SN

Whilst these additional principles UC and Vx go some way to further limiting the choice of w on SL in the absence of any prior knowledge it seems not immediately clear that they provide any particularly useful insights into the structure of w .

An altogether different approach to understanding binary induction is to first hazard an explanation for the nature of binary relations. Our initial example of Adam the gardener provides one such possible explanation. For while we have presented this as an example of a natural *binary* relation, ‘ X pollinates Y ’, between apple varieties there is another explanation which nowadays comes easily comes to mind (though presumable it would not have done to Adam had he lived in the 17th century) and which does not involve a direct binary relation between apple varieties at all. Namely, the reason variety A is a good pollinator and variety B is easily pollinated can be explained in terms of the (in this case genetic) properties that A and B have *as individuals*. In other words the binary relation of ‘ X pollinates Y ’ is actually determined not by some magic link between X and Y but by (unary) properties that X and Y possess in isolation. According to this picture then, whether or not a tree of variety A will be a successful pollinator of a tree of variety B is already determined by their individual properties *even before their paths cross*.⁹

Somewhat more precisely then this amounts to saying that ‘ X pollinates Y ’ is really equivalent to some, possibly very long, sentence involving only unary properties of X and Y ¹⁰. In [30] a construction based on this idea is given in which a probability function on a unary language with finitely many predicates is lifted to the binary language L by interpreting the binary relations $R_j(x, y)$ as random quantifier free sentences $\theta(x, y)$ of a (finite) unary language. Whilst we shall return to this idea again later what is of immediate interest about this construction as far as this paper is concerned is that investigation of the properties of the resulting probability functions on SL reveals that they satisfy a particular principle, the *Principle of Spectrum Exchangeability*, which is both natural and seems to provide a doorway towards understanding binary (and indeed higher arity) induction. We now explain this principle.

Given a state description for our binary language

$$\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j)$$

let $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ be the $p \times p$ matrix with ij -th entry r_{ij} and define the relation $I^{\mathbf{r}}$ to hold between $1 \leq i, j \leq p$ just if $r_{hj} = r_{hi}$ and $r_{jh} = r_{ih}$ for

⁹Rather in the way we can, in theory, by separately studying a key and a lock determine if they will fit without ever actually having to bring them together.

¹⁰There are at least two other positions which advocate explaining binary relations in terms of unary properties, see [33], but neither bears any resemblance to what is being proposed here.

$1 \leq h \leq p$. Let $\neg I^r$ be the complement of I^r . Clearly I^r is an equivalence relation on $\{1, 2, \dots, p\}$.

Define the *spectrum* of \mathbf{r} , $\mathcal{S}(\mathbf{r})$, to be the tuple $\langle |I_1|, |I_2|, \dots, |I_q| \rangle$ where $|I|$ is the number of elements of the set I and I_1, I_2, \dots, I_q are the equivalence classes of I^r arranged in non-increasing order of size.

The Spectrum Exchangeability Principle (Sx)

If $\mathbf{o}, \mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ have the same spectrum then

$$w \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right) = w \left(\bigwedge_{i,j=1}^p \beta_{o_{ij}}(a_i, a_j) \right).$$

We shall on occasions denote this left hand side expression by $w(\mathbf{r})$, or by $w(\vec{h})$ when w is known to satisfy Sx and $\mathcal{S}(\mathbf{r}) = \vec{h}$.

As we shall shortly show this principle implies Vx and UC. Firstly however it is of interest to note that the analogous property for the unary case is just the *Atom Exchangeability Principle*, Ax, that for v a probability function on a unary language with atoms $\alpha_1, \alpha_2, \dots, \alpha_{2^n}$ the value of v on the state description

$$\bigwedge_{i=1}^p \alpha_{h_i}(a_i)$$

is just a function of the *spectrum* of \vec{h} i.e. the vector of non-zero values $|\{i : h_i = j\}|$ arranged in non-increasing order¹¹. [For more details see for example [16].] This observation seems to us to point to the significance of this property and justifies us giving a name to it

Clearly the principles Px and SN (for this binary language) follow from Sx. The principle Ex also follows since for that it is enough to check that if $\mathbf{o} \in \{1, \dots, 2^m\}^{n \times n}$ is the result of transposing the i th and j th rows and columns in

¹¹The principle Ax has suffered considerable criticism on the grounds that, for example, if atoms α_1, α_2 are ‘close’ in the sense that they decide most of the (now unary) predicates of the language in the same way, i.e. true or false, then the observation that $\alpha_1(a_1)$ should provide more evidence for $\alpha_2(a_2)$ subsequently being observed than it would if α_1, α_2 were not close at all. (Attempting to broaden his continuum of inductive methods to encompass such a feature certainly occupied Carnap greatly in [3], though according to Maher, [26], [27], no attempt to date has been entirely successful.) There can be little doubt that in our everyday lives we often do subscribe to such a rule, since repetitions of experiences are rarely *identical* we really have to if we are to use the past to guide our future actions. One obvious justification for us doing this is that in practice most predicates are actually independent of each other, and furthermore we are aware of it. However it does not seem to us right to say that because this is *usually* the state of affairs then we should try to include that directly as a *principle* to be applied in the assumed absence of any prior knowledge. Indeed examples such as that given by Miller in [29] would seem to actively argue against taking this route.

$\mathbf{r} \in \{1, \dots, 2^m\}^{n \times n}$ then $w(\mathbf{o}) = w(\mathbf{r})$, and this indeed holds since $\mathcal{S}(\mathbf{o}) = \mathcal{S}(\mathbf{r})$. [This situation then parallels the unary case where Ex, Px, SN all follow from Ax.]

The main result of this section is the following theorem.

Theorem 1 *Sx implies Vx and UC.*

Proof We first show Vx. Suppose that $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ and that

$$\theta = \bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j).$$

Using the notation in the definition of Vx let

$$\theta' \equiv \bigwedge_{i,j=1}^p \beta_{o_{ij}}(a_i, a_j),$$

– notice that there is a unique such state description since θ , and hence θ' , decides $R_h(a_i, a_j)$ for all $1 \leq h \leq m$ and $1 \leq i, j \leq p$. Furthermore it is clear by considering cases that for $1 \leq i, j \leq p$,

$$I^{\mathbf{r}}(i, j) \iff I^{\mathbf{o}}(i, j)$$

so $\mathcal{S}(\mathbf{r}) = \mathcal{S}(\mathbf{o})$ and hence $w(\theta) = w(\theta')$ by Sx.

The result now follows for general θ as they are just equivalent to disjunctions of such state descriptions.

To prove that UC holds we first need to introduce some notation and prove a couple of lemmata. The purpose of these is to show that for $\vec{q} \in \{1, 2, \dots, 2^m\}^p$ and a given spectrum $\vec{s} \in \bigcup_{i=1}^{\infty} (\mathbb{Z}^+)^i$ the number of $p \times p$ matrices with entries from $\{1, 2, \dots, 2^m\}$, spectrum \vec{s} and k th column (or row) \vec{q} , where $1 \leq k \leq p$, is the same as the number of matrices with this same spectrum and diagonal \vec{q} . We do this by induction on p by pairing off matrices. So in the next lemma the idea is that \mathbf{r}, \mathbf{t} have been paired off at stage p and we go on to show how the matrices extending \mathbf{r}, \mathbf{t} to size $(p+1) \times (p+1)$ can likewise be successfully paired off.

Lemma 2 *Let $\mathbf{r}, \mathbf{t} \in \{1, 2, \dots, 2^m\}^{p \times p}$, $1 \leq q \leq 2^m$. Let I be an equivalence relation over $\{1, 2, \dots, p+1\}$ and let*

$$\mathcal{C}_q(I, \mathbf{r}) = \left\{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{(p+1) \times (p+1)} : \begin{array}{l} o_{ij} = r_{ij} \text{ for } 1 \leq i, j \leq p, \\ o_{1(p+1)} = q \\ I^{\mathbf{o}} = I \end{array} \right\},$$

$$\mathcal{D}_q(I, \mathbf{t}) = \left\{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{(p+1) \times (p+1)} : \begin{array}{l} o_{ij} = t_{ij} \text{ for } 1 \leq i, j \leq p, \\ o_{(p+1)(p+1)} = q \\ I^{\mathbf{o}} = I \end{array} \right\}.$$

Then if $r_{1j} = t_{jj}$ for $1 \leq j \leq p$ and $I^{\mathbf{r}} = I^{\mathbf{t}}$,

$$|\mathcal{C}_q(I, \mathbf{r})| = |\mathcal{D}_q(I, \mathbf{t})|.$$

Proof. If $I^{\mathbf{r}}$ and $I^{\mathbf{t}}$ have identical equivalence classes, say I_1, I_2, \dots, I_k , then for non-empty $\mathcal{C}_q(I, \mathbf{r})$ and $\mathcal{D}_q(I, \mathbf{t})$, I must have equivalence classes equivalent, modulo the removal of empty sets, to one of the following three forms of partition of $\{1, 2, \dots, p+1\}$:

- (a) $\{I_1, I_2, \dots, I_{l-1}, I_l \cup \{p+1\}, I_{l+1}, \dots, I_k\}$ for some $1 \leq l \leq k$,
- (b) $\{I_1, I_2, \dots, I_k, \{p+1\}\}$,
- (c) $\{J_{11}, J_{12}, \dots, J_{12^{2^m}}, J_{21}, J_{22}, \dots, J_{22^{2^m}}, \dots, J_{k1}, J_{k2}, \dots, J_{k2^{2^m}}, \{p+1\}\}$,

where the sets $\{J_{i1}, J_{i2}, \dots, J_{i2^{2^m}}\}$ partition I_i for $1 \leq i \leq k$ and at least one of these is a non-trivial partition i.e. case (b) does not apply.

Suppose I is of the form (a). Then

$$\begin{aligned} |\mathcal{C}_q(I, \mathbf{r})| &= \begin{cases} 0 & \text{if } q \neq r_{1j} \text{ for } 1 \leq j \leq p \text{ such that } j \in I_l \\ 1 & \text{otherwise.} \end{cases} \\ |\mathcal{D}_q(I, \mathbf{t})| &= \begin{cases} 0 & \text{if } q \neq t_{jj} \text{ for } 1 \leq j \leq p \text{ such that } j \in I_l, \\ 1 & \text{otherwise.} \end{cases} \end{aligned}$$

and since $r_{1j} = t_{jj}$ for $1 \leq j \leq p$, we obtain

$$|\mathcal{C}_q(I, \mathbf{r})| = |\mathcal{D}_q(I, \mathbf{t})|.$$

Suppose I is of the form (c). For $\mathbf{o} \in \mathcal{C}_q(I, \mathbf{r})$ it holds by definition of $\mathcal{C}_q(I, \mathbf{r})$ that

$$\begin{aligned} o_{ij} &= r_{ij} && \text{for } 1 \leq i, j \leq p, \\ o_{1(p+1)} &= q, \\ 1 \leq o_{(i+1)(p+1)}, o_{(p+1)j} &\leq 2^m && \text{for } 1 \leq i, j \leq p. \end{aligned}$$

Similarly, for $\mathbf{o} \in \mathcal{D}_q(I, \mathbf{t})$ it holds that

$$\begin{aligned} o_{ij} &= t_{ij} && \text{for } 1 \leq i, j \leq p, \\ o_{(p+1)(p+1)} &= q, \\ 1 \leq o_{i(p+1)}, o_{(p+1)j} &\leq 2^m && \text{for } 1 \leq i, j \leq p. \end{aligned}$$

In both cases we have for $1 \leq l \leq k$, $1 \leq h, g \leq 2^{2^m}$ and $i \in J_{lh}$, $j \in J_{lg}$

$$\begin{aligned} \langle o_{j(p+1)}, o_{(p+1)j} \rangle &= \langle o_{i(p+1)}, o_{(p+1)i} \rangle && \text{if } h = g, \\ \langle o_{j(p+1)}, o_{(p+1)j} \rangle &\neq \langle o_{i(p+1)}, o_{(p+1)i} \rangle && \text{if } h \neq g. \end{aligned}$$

So, if for $1 \leq l \leq k$ there are c_l non-empty $J_{l1}, J_{l2}, \dots, J_{l2^{2^m}}$ then we see there are

$$\prod_{l=1}^k \frac{2^{2^m}!}{(2^{2^m} - c_l)!}$$

distinct $\mathbf{o} \in \mathcal{D}_q(I, \mathbf{t})$ since if we assume without loss of generality for each $1 \leq l \leq k$ that the non-empty J_{lh} for $1 \leq h \leq 2^{2m}$ are $J_{l1}, J_{l2}, \dots, J_{lc_l}$ then we have 2^{2m} choices for the pair $\langle o_{i(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{l1}$, then $2^{2m} - 1$ choices for the pair $\langle o_{i(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{l2}$, and so on, until we have $2^{2m} - c_l + 1$ choices for the pair $\langle o_{i(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{lc_l}$. Notice that since this is a non-trivial partition for all these choices $p + 1$ will end up in a equivalence class by itself.

Similarly the number of distinct $\mathbf{o} \in \mathcal{C}_q(I, \mathbf{r})$ is also

$$\prod_{l=1}^k \frac{2^{2m}!}{(2^{2m} - c_l)!}$$

for again if we assume without loss of generality for each $1 \leq l \leq k$ that the non-empty J_{lh} for $1 \leq h \leq 2^{2m}$ are $J_{l1}, J_{l2}, \dots, J_{lc_l}$ and that $1 \in J_{l1}$ then this time we have 2^{2m} choices for the pair $\langle o_{(p+1)(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{l1}$ and 2^{2m} choices for all other $\langle o_{i(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{l1}$ for $2 \leq l \leq k$, and continuing as before for each $1 \leq l \leq k$ we have $2^{2m} - 1$ choices for the pair $\langle o_{i(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{l2}$, and so on, until we have $2^{2m} - c_l + 1$ choices for the pair $\langle o_{i(p+1)}, o_{(p+1)i} \rangle$ where $i \in J_{lc_l}$.

Finally suppose that (b) applies. If we follow the same path as in case (c) everything works out fine *except* that for some of the choices $p + 1$ will join one of the old equivalences classes rather than forming a new equivalence class by itself. In the case of \mathbf{r} this will happen once for each each equivalence class I_l for which $r_{1j} = q$ for $j \in I_l$, so these choices need to be removed from the grand product obtained as in (c). Similarly in the case of \mathbf{o} we need to remove one choice for each equivalence class I_l such that $o_{jj} = q$ for $j \in I_l$. However since $r_{1j} = t_{jj}$ for $j = 1, 2, \dots, p$ the number of removals will be the same in both cases so again we have equality. \blacksquare

Lemma 3 For $\vec{q} \in \{1, 2, \dots, 2^m\}^p$ and $\vec{s} \in \bigcup_{i=1}^{\infty} (\mathbb{Z}^+)^i$ let

$$\begin{aligned} \mathcal{A}^{k,p}(\vec{q}, \vec{s}) &= \{\mathbf{o} \in \{1, 2, \dots, 2^m\}^{p \times p} : \mathcal{S}(\mathbf{o}) = \vec{s}, o_{ik} = q_i \text{ for } 1 \leq i \leq p\} \\ \mathcal{A}_{k,p}(\vec{q}, \vec{s}) &= \{\mathbf{o} \in \{1, 2, \dots, 2^m\}^{p \times p} : \mathcal{S}(\mathbf{o}) = \vec{s}, o_{ki} = q_i \text{ for } 1 \leq i \leq p\} \end{aligned}$$

for $1 \leq k \leq p$ and let

$$\mathcal{B}_p(\vec{q}, \vec{s}) = \{\mathbf{o} \in \{1, 2, \dots, 2^m\}^{p \times p} : \mathcal{S}(\mathbf{o}) = \vec{s}, o_{ii} = q_i \text{ for } 1 \leq i \leq p\}.$$

Then

$$|\mathcal{B}_p(\vec{q}, \vec{s})| = |\mathcal{A}^{k,p}(\vec{q}, \vec{s})| = |\mathcal{A}_{k,p}(\vec{q}, \vec{s})| = |\mathcal{A}_{l,p}(\vec{q}, \vec{s})|$$

for any $1 \leq k, l \leq p$.

Proof. That $\mathcal{A}_{k,p}(\vec{q}, \vec{s}) = \mathcal{A}_{l,p}(\vec{q}, \vec{s})$ for $1 \leq k, l \leq p$ is clear by transposing the k th and l th columns and rows in the matrices in $\mathcal{A}_{k,p}(\vec{q}, \vec{s})$. Similarly it follows

that $\mathcal{A}_{k,p}(\vec{q}, \vec{s}) = \mathcal{A}^{k,p}(\vec{q}, \vec{s})$ for $1 \leq k \leq p$ by taking the transposes of the matrices in $\mathcal{A}_{k,p}(\vec{q}, \vec{s})$. So it suffices to prove that $|\mathcal{B}_p(\vec{q}, \vec{s})| = |\mathcal{A}_{1,p}(\vec{q}, \vec{s})|$.

By Lemma 2 we can iteratively form a one to one correspondence between the sequences $\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_p$ such that $\mathbf{r}_l \in \{1, 2, \dots, 2^m\}^{l \times l}$ for $0 \leq l \leq p$ with the property $\mathbf{r}_l \in \mathcal{C}_{q_l}(I^{\mathbf{r}_l}, \mathbf{r}_{l-1})$ for $1 \leq l \leq p$ and the sequences $\mathbf{t}_0, \mathbf{t}_1, \dots, \mathbf{t}_p$ such that $\mathbf{t}_l \in \{1, 2, \dots, 2^m\}^{l \times l}$ for $0 \leq l \leq p$ with the property $\mathbf{t}_l \in \mathcal{D}_{q_l}(I^{\mathbf{t}_l}, \mathbf{t}_{l-1})$ in such a way that $I^{\mathbf{r}_l} = I^{\mathbf{t}_l}$ for $1 \leq l \leq p$. Hence, since \mathbf{r}_p determines $\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_p$ and \mathbf{t}_p determines $\mathbf{t}_0, \mathbf{t}_1, \dots, \mathbf{t}_p$ we have a one to one correspondence confirming that

$$\begin{aligned} & |\{\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p} : r_{1i} = q_i \text{ for } 1 \leq i \leq p, I^{\mathbf{r}} = I^{\mathbf{o}}\}| \\ &= |\{\mathbf{t} \in \{1, 2, \dots, 2^m\}^{p \times p} : t_{ii} = q_i \text{ for } 1 \leq i \leq p, I^{\mathbf{t}} = I^{\mathbf{o}}\}| \end{aligned}$$

whenever $\mathbf{o} \in \{1, 2, \dots, 2^m\}^{p \times p}$. That $|\mathcal{B}_p(\vec{q}, \vec{s})| = |\mathcal{A}_{1,p}(\vec{q}, \vec{s})|$ now follows by considering the disjoint union of our injective maps over all equivalence relations $I^{\mathbf{o}}$ for $\mathbf{o} \in \{1, 2, \dots, 2^m\}^{p \times p}$ such that $\mathcal{S}(\mathbf{o}) = \vec{s}$. \blacksquare

Proof of Theorem 1 continued.

First let $1 \leq k \leq p$ and let w satisfy Sx, so without ambiguity we can introduce the shorthand notation

$$w \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right) = w(\vec{s})$$

where $\vec{s} = \mathcal{S}(\mathbf{r})$. Then

$$\begin{aligned} w \left(\bigwedge_{i=1}^p \beta_{q_i}(a_i, a_k) \right) &= \sum_{\substack{\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p} \\ r_{ik} = q_i \text{ for } 1 \leq i \leq p}} w(\mathbf{r}) \\ &= \sum_{\vec{s} \in \bigcup_{i=1}^{\infty} (\mathbf{Z}^+)^i} |\mathcal{A}^{k,p}(\vec{q}, \vec{s})| w(\vec{s}) \end{aligned} \quad (1)$$

and

$$\begin{aligned} w \left(\bigwedge_{i=1}^p \beta_{q_i}(a_i, a_i) \right) &= \sum_{\substack{\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p} \\ r_{ii} = q_i \text{ for } 1 \leq i \leq p}} w(\mathbf{r}) \\ &= \sum_{\vec{s} \in \bigcup_{i=1}^{\infty} (\mathbf{Z}^+)^i} |\mathcal{B}_p(\vec{q}, \vec{s})| w(\vec{s}) \end{aligned} \quad (2)$$

and so $w(\bigwedge_{i=1}^p \beta_{q_i}(a_i, a_k)) = w(\bigwedge_{i=1}^p \beta_{q_i}(a_i, a_i))$ by Lemma 3. Similarly for each other pair of equations in UC.

Finally for $p < k$ the required equality follows by taking (1) and (2) with k in place of p and summing over all choices of $q_{p+1}, q_{p+2}, \dots, q_k$. \blacksquare

A Representation Theorem

In this section we shall first establish a representation theorem for a rather special class of binary probability functions satisfying \mathcal{S}_x and then widen its scope. First however we need some useful notation, not least to actually describe this special class.

Let

$$A_p^t = \{ \langle p_1, p_2, \dots, p_t \rangle : p_1 \geq p_2 \geq \dots \geq p_t > 0 \text{ and } \sum_i p_i = p \}$$

and

$$A_p = \bigcup_t A_p^t.$$

Given $\vec{p} \in A_p$ and $\vec{q} \in A_q$ let $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ be such that $\mathcal{S}(\mathbf{r}) = \vec{p}$ and define $\mathcal{N}(\vec{p}, \vec{q})$ to be

$$|\{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q} : \mathcal{S}(\mathbf{o}) = \vec{q}, \text{ and } o_{ij} = r_{ij} \text{ for } 1 \leq i, j \leq p \}|.$$

In other words $\mathcal{N}(\vec{p}, \vec{q})$ is the number of $q \times q$ matrices with spectrum \vec{q} which extend some particular $p \times p$ matrix \mathbf{r} with spectrum \vec{p} . The fact that $\mathcal{N}(\vec{p}, \vec{q})$ is well defined, i.e. that it does not depend on the particular choice of \mathbf{r} , is guaranteed by the next Lemma.

Lemma 4 *If $\mathbf{r}, \mathbf{u} \in \{1, 2, \dots, 2^m\}^{p \times p}$ are such that $\mathcal{S}(\mathbf{r}) = \mathcal{S}(\mathbf{u}) = \vec{p}$ then the two sets*

$$\{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q} : \mathcal{S}(\mathbf{o}) = \vec{q}, \text{ and } o_{ij} = r_{ij} \text{ for } 1 \leq i, j \leq p \}$$

$$\{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q} : \mathcal{S}(\mathbf{o}) = \vec{q}, \text{ and } o_{ij} = u_{ij} \text{ for } 1 \leq i, j \leq p \}$$

have the same number of elements.

We omit the proof of this Lemma because it essentially uses the analysis in the proof of Lemma 2 to count the number of such \mathbf{o} in the case $q = p + 1$ and observe that the answer depends only on \vec{p} , the general case following by iterating this process.

Similarly in the unary case, given $\vec{h} \in \{1, 2, \dots, 2^n\}^p$ with spectrum \vec{p} let $\mathcal{F}(\vec{p}, \vec{q})$ be

$$|\{ \vec{g} \in \{1, 2, \dots, 2^n\}^q : g_i = h_i \text{ for } 1 \leq i \leq p \text{ and } \mathcal{S}(\vec{g}) = \vec{q} \}|.$$

Again this is easily seen to depend only on \vec{p} and not on the particular choice of \vec{h} .

Lemma 5 *Let $s \leq 2^n$. Then for some non-zero constant $d(s)$ depending only on the length s of the spectrum \vec{p} ,*

$$\mathcal{N}(\emptyset, \vec{p}) = d(s)\mathcal{F}(\emptyset, \vec{p})$$

where \emptyset here denotes the 0-vector.

Proof. Consider a partition B_1, B_2, \dots, B_s of $\{1, 2, \dots, p\}$ into non-empty sets listed in such a way that 1 is the least element of B_1 and in general the least element of B_j is also the least element of

$$\{1, 2, \dots, p\} - \bigcup_{r < j} B_r. \quad (3)$$

Let $\mathbf{t} \in \{1, 2, \dots, 2^m\}^{s \times s}$ have spectrum of length s (i.e. spectrum $\vec{1}_s$). Then B_1, B_2, \dots, B_s and \mathbf{t} uniquely determine an $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ with equivalence classes B_1, B_2, \dots, B_s by setting

$$r_{ij} = t_{g(i)g(j)}, \text{ where } i \in B_{g(i)}, j \in B_{g(j)}.$$

Furthermore every matrix $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ with spectrum of length s determines a unique such $\mathbf{t} \in \{1, 2, \dots, 2^m\}^{s \times s}$ and equivalence classes B_1, \dots, B_s satisfying (3). For future reference we shall refer to $\mathbf{t}, B_1, \dots, B_s$ as the *canonical representation of \mathbf{r}* .

It follows then that for a fixed such B_1, \dots, B_s the number of such \mathbf{r} is the number, $T_s (\neq 0)$, of such \mathbf{t} .

Exactly similarly the number of $\vec{h} \in \{1, 2, \dots, 2^n\}^p$ which yield the equivalence classes B_1, B_2, \dots, B_s ordered as above (i.e. satisfying (3)) is the (non-zero) number, U_s , of (ordered) choices of $\alpha_{h_{\min B_1}}, \alpha_{h_{\min B_2}}, \dots, \alpha_{h_{\min B_s}}$. Since T_s, U_s depend only on s summing over all such B_1, B_2, \dots, B_s which yield spectrum \vec{p} gives the result. \blacksquare

Let w satisfy Sx. We say that w is $\leq t$ -heterogeneous¹² if $w(\vec{s}) = 0$ whenever the length of spectrum \vec{s} , denoted by $|\vec{s}|$, exceeds t . We say that w is t -heterogeneous if in addition

$$\lim_{k \rightarrow \infty} \sum_{\vec{k} \in \bigcup_{s < t} A_k^s} \mathcal{N}(\emptyset, \vec{k})w(\vec{k}) = 0. \quad (4)$$

In other words w is t -heterogeneous if in the limit all the probability is massed on the spectra of length exactly t . Since for spectrum \vec{h} with $k \geq \sum h_i$,

$$w(\vec{h}) = \sum_{\vec{k} \in A_k} \mathcal{N}(\vec{h}, \vec{k})w(\vec{k})$$

¹²What is here being called $\leq t$ -heterogeneous was called just t -heterogeneous in [30].

this means that

$$w(\vec{h}) = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^t} \mathcal{N}(\vec{h}, \vec{k}) w(\vec{k}). \quad (5)$$

In contrast to heterogeneity we say that w satisfying S_x is *homogeneous* if for all s ,

$$\lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^s} \mathcal{N}(\emptyset, \vec{k}) w(\vec{k}) = 0. \quad (6)$$

Our interest in $\leq t$ -heterogeneous probability functions arises in particular because the binary probability functions alluded to earlier which result, as described in [30], from identifying binary relations with sentences in a language with a fixed finite number n of unary predicates have the property of being $\leq 2^n$ -heterogeneous.

We shall adopt this same notation for probability functions v on a unary language (with n predicates as usual) satisfying Atom Exchangeability, A_x . Namely, we shall write $v(\vec{s})$ for

$$v \left(\bigwedge_{i=1}^m \alpha_{h_i}(a_i) \right)$$

when \vec{s} is the spectrum, $\mathcal{S}(\vec{h})$, of \vec{h} and say that v is $\leq t$ -heterogeneous if $v(\vec{s}) = 0$ whenever $|\vec{s}| > t$, and t -heterogeneous when in addition

$$\lim_{k \rightarrow \infty} \sum_{\vec{k} \in \bigcup_{s < t} A_k^s} \mathcal{F}(\emptyset, \vec{k}) v(\vec{k}) = 0.$$

Again in this case

$$v(\vec{h}) = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^t} \mathcal{F}(\vec{h}, \vec{k}) v(\vec{k}). \quad (7)$$

We now prove a representation theorem for t -heterogeneous binary probability functions which shows them to be closely related to t -heterogeneous unary probability functions¹³.

Theorem 6 *Let w be a binary t -heterogeneous probability function on a language with m binary predicates and satisfying S_x . Let $2^n \geq t$ and define a function v on a unary language with n unary predicates as follows. For $\vec{g} \in \{1, 2, \dots, 2^n\}^p$ with $\mathcal{S}(\vec{g}) = \vec{h}$ define*

$$v \left(\bigwedge_{i=1}^p \alpha_{g_i}(a_i) \right) = v(\vec{h}) = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^t} d(t) \mathcal{F}(\vec{h}, \vec{k}) w(\vec{k}). \quad (8)$$

¹³An alternative Representation Theorem is given in [30].

Then v is (more precisely, extends to) a t -heterogeneous unary probability function satisfying Ax . Conversely if v is a t -heterogeneous unary probability function on this language satisfying Ax and we define function w on the state descriptions of a binary language with m predicates by

$$w\left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j)\right) = w(\vec{h}) = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^t} d(t)^{-1} \mathcal{N}(\vec{h}, \vec{k}) v(\vec{k}) \quad (9)$$

when $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ and $\mathcal{S}(\mathbf{r}) = \vec{h}$, then w is (more precisely, extends to) a t -heterogeneous binary probability function satisfying Sx . Furthermore these two operations are inverses to each other.

Proof. First assume that the limit in (8) exists. To show that v determines a probability function satisfying Ax it is enough that

$$v(\emptyset) = 1, \quad (10)$$

$$v(\vec{h}) = \sum_{\vec{q} \in A_{h+1}} \mathcal{F}(\vec{h}, \vec{q}) v(\vec{q}), \quad (11)$$

where $\vec{h} \in A_h$.

That (10) holds is clear from (5) and the definition of $d(t)$. To show (11) it is enough to notice that for $\sum k_i > h$,

$$\mathcal{F}(\vec{h}, \vec{k}) = \sum_{\vec{q} \in A_{h+1}} \mathcal{F}(\vec{h}, \vec{q}) \mathcal{F}(\vec{q}, \vec{k}).$$

To show that the limit exists notice that for $\vec{k} \in A_k^t$, $\vec{p} \in A_p^t$, with $k < p$,

$$\mathcal{F}(\vec{k}, \vec{p}) = \mathcal{N}(\vec{k}, \vec{p}) \quad (12)$$

and

$$w(\vec{k}) = \sum_{\vec{p} \in A_p^t} \mathcal{N}(\vec{k}, \vec{p}) w(\vec{p}).$$

Hence

$$\begin{aligned} \sum_{\vec{k} \in A_k^t} d(t) \mathcal{F}(\vec{h}, \vec{k}) w(\vec{k}) &= \sum_{\vec{k} \in A_k^t} \sum_{\vec{p} \in A_p^t} d(t) \mathcal{F}(\vec{h}, \vec{k}) \mathcal{N}(\vec{k}, \vec{p}) w(\vec{p}) \\ &= \sum_{\vec{k} \in A_k^t} \sum_{\vec{p} \in A_p^t} d(t) \mathcal{F}(\vec{h}, \vec{k}) \mathcal{F}(\vec{k}, \vec{p}) w(\vec{p}) \text{ by (12)} \\ &\leq \sum_{\vec{p} \in A_p^t} d(t) \mathcal{F}(\vec{h}, \vec{p}) w(\vec{p}) \end{aligned}$$

since for $k < p, \vec{p} \in A_p^t$,

$$\sum_{\vec{k} \in A_k^t} \mathcal{F}(\vec{h}, \vec{k}) \mathcal{F}(\vec{k}, \vec{p}) \leq \mathcal{F}(\vec{h}, \vec{p}).$$

It follows then that the terms in the sequence defining $v(\vec{h})$ are increasing. They are also bounded since

$$\sum_{\vec{k} \in A_k^t} d(t) \mathcal{F}(\vec{h}, \vec{k}) w(\vec{k}) \leq \sum_{\vec{k} \in A_k^t} d(t) \mathcal{F}(\emptyset, \vec{k}) w(\vec{k}) = \sum_{\vec{k} \in A_k^t} \mathcal{N}(\emptyset, \vec{k}) w(\vec{k}) = 1.$$

Hence the required limit exists.

Finally v is t -heterogeneous since for $\vec{k} \in A_k^t$,

$$\mathcal{F}(\emptyset, \vec{k}) v(\vec{k}) = \mathcal{N}(\emptyset, \vec{k}) w(\vec{k}),$$

so for $s < t$

$$\sum_{\vec{k} \in A_k^s} \mathcal{F}(\emptyset, \vec{k}) v(\vec{k}) \leq 1 - \sum_{\vec{k} \in A_k^t} \mathcal{F}(\emptyset, \vec{k}) v(\vec{k}) = 1 - \sum_{\vec{k} \in A_k^t} \mathcal{N}(\emptyset, \vec{k}) w(\vec{k})$$

and because w is t -heterogeneous this right hand term tends to zero as k tends to ∞ .

In an exactly similar fashion we can show that if conversely the unary v (on a unary language with n predicates) satisfies Ax and is t -heterogeneous then the function w defined on the binary language with m predicates by (9) is a well defined (i.e. the limit exists) t -heterogeneous binary probability function satisfying Sx.

To conclude the proof we need to show that these operations are inverses to each other. To see this notice that since

$$w(\vec{p}) = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^t} \mathcal{N}(\vec{p}, \vec{k}) w(\vec{k})$$

for $\vec{p} \in A_p$ it is enough to show that for $p < k < q$

$$\sum_{\vec{k} \in A_k^t} d(t)^{-1} \mathcal{N}(\vec{p}, \vec{k}) \cdot \left\{ \sum_{\vec{q} \in A_q^t} d(t) \mathcal{F}(\vec{k}, \vec{q}) w(\vec{q}) \right\} = \sum_{\vec{k} \in A_k^t} \mathcal{N}(\vec{p}, \vec{k}) w(\vec{k}). \quad (13)$$

But the expression on the left hand side of (13) is just

$$\sum_{\vec{k} \in A_k^t} \sum_{\vec{q} \in A_q^t} \mathcal{N}(\vec{p}, \vec{k}) \mathcal{F}(\vec{k}, \vec{q}) w(\vec{q})$$

and since

$$\mathcal{F}(\vec{k}, \vec{q}) = \mathcal{N}(\vec{k}, \vec{q})$$

this gives the required equality. Showing the ‘inverse relationship’ in the other direction is similar. ■

We can to some extent generalize the above results as follows. Let w be a binary probability function (on the language with m binary predicates as usual) satisfying Sx. Define

$$\gamma_1 = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^1} \mathcal{N}(\emptyset, \vec{k}) w(\vec{k}).$$

Clearly this limit exists since the terms are non-increasing (and bounded below by zero). Assuming that $\gamma_1 \neq 0$ define

$$w^{[1]} \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_{t_i}, a_{t_j}) \right) = w^{[1]}(\vec{h}) = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^1} \gamma_1^{-1} \mathcal{N}(\vec{h}, \vec{k}) w(\vec{k})$$

where $\vec{h} = \mathcal{S}(\mathbf{r})$. Similarly to the proof of the previous result it is straightforward to check that provided $\gamma_1 \neq 0$ $w^{[1]}$ is a 1-heterogeneous probability function. Assume for the moment $\gamma_1 < 1$ and set

$$w^{(1)} = \begin{cases} (1 - \gamma_1)^{-1} (w - \gamma_1 w^{[1]}) & \text{if } \gamma_1 \neq 0, \\ w & \text{if } \gamma_1 = 0. \end{cases}$$

Then $w^{(1)}$ is a probability function (satisfying Sx),

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

(where we just take this first right hand term to be zero if $\gamma_1 = 0$) and

$$\lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^1} \mathcal{N}(\emptyset, \vec{k}) w^{(1)}(\vec{k}) = 0.$$

In a similar fashion, with 2 in place of 1 and starting with $w^{(1)}$ in place of w we can form $\gamma_2, w^{[2]}, w^{(2)}$ so that,

$$\gamma_2 = \lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^2} \mathcal{N}(\emptyset, \vec{k}) w^{(1)}(\vec{k}).$$

and assuming $0 < \gamma_2 < 1$,

$$w^{(1)} = \gamma_2 w^{[2]} + (1 - \gamma_2) w^{(2)}$$

with $w^{[2]}$ 2-heterogeneous and

$$\lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^i, i \leq 2} \mathcal{N}(\emptyset, \vec{k}) w^{(2)}(\vec{k}) = 0.$$

The pattern should now be clear. Continuing in this way gives the following result.

Theorem 7 *Let w be a binary probability function (on the language with m binary predicates) satisfying Sx . Then there are binary probability functions $w^{[t]}$ (satisfying Sx) and constants $\eta_t \geq 0$ for $0 \leq t < \infty$ such that*

$$w = \sum_{i=0}^{\infty} \eta_i w^{[i]}, \quad \sum_{i=0}^{\infty} \eta_i = 1,$$

$w^{[t]}$ is t -heterogeneous for $t > 0$ and $w^{[0]}$ is homogeneous (see (6)). Furthermore the η_i are unique and so are the $w^{[i]}$ when $\eta_i \neq 0$.

We conclude this section by giving another representation theorem for the t -heterogeneous case which provides a sort of converse to the original construction of binary probability functions alluded to in the second section of this paper where the $R_1(x, y), \dots, R_m(x, y)$ are identified with a (uniform) random vector of quantifier free 2-ary formulae (up to logical equivalence), $\psi_1(x, y), \dots, \psi_m(x, y)$, from a unary language with n predicates. Such an identification leads from a probability function on the unary language (satisfying Ex) to one on the binary language satisfying Sx. In addition this probability function on the binary language turns out to be $\leq 2^n$ -heterogeneous.

What we shall now show is that *all* $\leq 2^n$ -heterogeneous binary probability functions arise in this way *provided* we only require that v satisfies Ex and replace ‘uniform’ by a possibly non-uniform distribution. In other words, make certain vectors of 2-ary formulae from the unary language more likely than others to represent the vector of binary predicates.

Before proceeding further however we need to explain a little more carefully the original construction in [30]. The precise idea there is of identifying the binary relations $R_1(x, y), R_2(x, y), \dots, R_m(x, y)$ with a (uniform) random vector of formulae of the form

$$\bigvee_{\langle \alpha_i, \alpha_j \rangle \in X} (\alpha_i(x) \wedge \alpha_j(y))$$

where the X run through all sets of ordered pairs of atoms from the unary language with predicates P_1, \dots, P_n . Notice that every 2-ary formula from this language is logically equivalent to such a formula for exactly one such X . This

explains our early parenthetic ‘up to logical equivalence’. Via this identification and a probability function v on this unary language satisfying Ex we obtain a probability function w on the binary language satisfying Sx as

$$w\left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_{k_i}, a_{k_j})\right) = \sum_{X_1, \dots, X_m} \rho(X_1, X_2, \dots, X_m) v\left(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_{k_i}, a_{k_j})\right)$$

where ρ is just the uniform distribution (so here $\rho(X_1, X_2, \dots, X_m) = 2^{-m2^{2n}}$) and $\psi_{r_{ij}}(\vec{X})(x, y)$ is just the result of replacing each $R_h(x, y)$ in $\beta_{r_{ij}}$ by

$$\bigvee_{\langle \alpha, \alpha' \rangle \in X_h} (\alpha(x) \wedge \alpha'(y)).$$

The next theorem tells us that any $\leq 2^n$ -heterogeneous binary probability function w satisfying Sx is of this form *provided* we allow that ρ may be non-uniform and that v need only satisfy Ex.

Theorem 8 *Let w be a $\leq 2^n$ -heterogeneous binary probability function satisfying Sx. Then, with the above notation, there is a probability function v on a unary language (satisfying Ex) and a distribution ρ such that*

$$w\left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_{k_i}, a_{k_j})\right) = \sum_{X_1, \dots, X_m} \rho(X_1, X_2, \dots, X_m) v\left(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_{k_i}, a_{k_j})\right).$$

Proof. We first prove the result when w is 2^n -heterogeneous. Define the 2^n -heterogeneous unary probability function v on a unary language with n predicates (P_1, \dots, P_n , as usual) according to (8), so w can be recovered from v by (9). So for $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ with spectrum \vec{h} of length 2^n ,

$$w(\mathbf{r}) = d(2^n)^{-1} v(\vec{h}). \quad (14)$$

We shall show that the theorem holds for this v and a suitable choice of ρ .

Without loss of generality assume that the atoms with respect to P_1, \dots, P_n are enumerated as $\alpha_1, \alpha_2, \dots, \alpha_{2^n}$ and that for $c - 1 = c_1 c_2 c_3 \dots c_m$ in binary,

$$\beta_c(x, y) = \bigwedge_{l=1}^m R_l^{c_l}(x, y).$$

Similarly for X_1, X_2, \dots, X_m a vector of sets of pairs of atoms $\langle \alpha_i, \alpha_j \rangle$ let \mathbf{x} be the $2^n \times 2^n \{1, 2, \dots, 2^m\}$ -matrix with

$$\mathbf{x}_{ij} = \sum_{l=1}^m c_l 2^{l-1}$$

where

$$c_l = \begin{cases} 1 & \text{if } \langle \alpha_i, \alpha_j \rangle \in X_l, \\ 0 & \text{otherwise.} \end{cases}$$

A convenient consequence of this choice of indexing is that for $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ the term $v(\alpha_{g_1}(a_1) \wedge \alpha_{g_2}(a_2) \wedge \dots \wedge \alpha_{g_p}(a_p))$ appears in the state description expansion of $v(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_i, a_j))$ just if

$$r_{ij} = x_{g_i g_j} \text{ for } i, j = 1, \dots, p. \quad (15)$$

Let

$$\rho(\vec{X}) = \begin{cases} |\{ \mathbf{o} \in \{1, \dots, 2^m\}^{2^n \times 2^n} : |\mathcal{S}(\mathbf{o})| = 2^n \}|^{-1} & \text{if } |\mathcal{S}(\mathbf{x})| = 2^n, \\ 0 & \text{otherwise.} \end{cases}$$

In other words $\rho(\vec{X}) = \mathcal{N}(\emptyset, \vec{1}_{2^n})^{-1}$ when $|\mathcal{S}(\mathbf{x})| = 2^n$ and zero otherwise, where $\vec{1}_{2^n}$ is the vector of length 2^n and all entries 1.

We shall show that for $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ with spectrum \vec{h} of length 2^n ,

$$w(\mathbf{r}) = \sum_{\vec{X}} \rho(\vec{X}) v\left(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_i, a_j)\right). \quad (16)$$

To see this notice that if

$$v(\alpha_{g_1}(a_1) \wedge \alpha_{g_2}(a_2) \wedge \dots \wedge \alpha_{g_p}(a_p)) \quad (17)$$

appears in the state description expansion of

$$v\left(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_i, a_j)\right) \quad (18)$$

then 2^n of the g_l must be distinct, otherwise there would be i, j such that $\neg I^{\mathbf{r}}(i, j)$ but $g_i = g_j$, which forces $I^{\mathbf{r}}(i, j)$ by (15).

Indeed by (15) we see that when $\rho(\vec{X}) \neq 0$ we can re-write (17) as

$$v\left(\bigwedge_{j=1}^{h_1} \alpha_{e_1}(a_{s_{1j}}) \wedge \bigwedge_{j=1}^{h_2} \alpha_{e_2}(a_{s_{2j}}) \wedge \dots \wedge \bigwedge_{j=1}^{h_{2^n}} \alpha_{e_{2^n}}(a_{s_{2^n j}})\right) \quad (19)$$

where the $\{s_{f1}, s_{f2}, \dots, s_{fh_f}\}$, for $f = 1, 2, \dots, 2^n$, are the equivalence classes of $I^{\mathbf{r}}$, ordered according to the canonical representation, and the e_1, e_2, \dots, e_{2^n} run through the 2^n different members of g_1, g_2, \dots, g_p . For fixed e_1, e_2, \dots, e_{2^n} (in that order) and \vec{X} with $\rho(\vec{X}) \neq 0$ the term (19) will appear in the expansion just if

$$x_{e_i e_j} = r_{s_{i1} s_{j1}} \text{ for } i, j = 1, 2, \dots, 2^n. \quad (20)$$

Clearly there is only one \vec{X} with this property. Altogether then, summing over the $2^n!$ choices for e_1, e_2, \dots, e_{2^n} , the coefficient of $v(\vec{h})$ (including $\rho(\vec{X})$) in the state description expansion of (16) is

$$2^n! \mathcal{N}(\emptyset, \vec{1}_{2^n})^{-1} = \mathcal{F}(\emptyset, \vec{1}_{2^n}) / \mathcal{N}(\emptyset, \vec{1}_{2^n}) = d(2^n)^{-1}.$$

Using this (16) follows from (14) in the case of \mathbf{r} with spectrum of length 2^n .

To show that it holds for general \mathbf{r} define the probability function w' by

$$w' \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right) = \sum_{\vec{X}} \rho(\vec{X}) v \left(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_i, a_j) \right). \quad (21)$$

Then for $k > p$,

$$\begin{aligned} w' \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right) &= \sum_{\mathbf{o} \in D_k} w' \left(\bigwedge_{i,j=1}^k \beta_{o_{ij}}(a_i, a_j) \right) \\ &\quad + \sum_{\mathbf{o} \in E_k} w' \left(\bigwedge_{i,j=1}^k \beta_{o_{ij}}(a_i, a_j) \right) \end{aligned}$$

where

$$D_k = \{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{k \times k} : |\mathcal{S}(\mathbf{o})| = 2^n \text{ and } o_{ij} = r_{ij} \text{ for } i, j = 1, \dots, p \},$$

$$E_k = \{ \mathbf{o} \in \{1, 2, \dots, 2^m\}^{k \times k} : |\mathcal{S}(\mathbf{o})| \neq 2^n \text{ and } o_{ij} = r_{ij} \text{ for } i, j = 1, \dots, p \}.$$

Since w and w' agree on the arguments in the first sum and w is 2^n -heterogeneous this first sum tends to

$$w \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right)$$

as $k \rightarrow \infty$ so

$$w' \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right) \geq w \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right). \quad (22)$$

Indeed since the sum of the right hand side of (22) over *all* $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ must be one we can conclude that in fact equality holds in (22), as required.

We now generalize this result to w t -heterogeneous for $t < 2^n$. In this case the unary probability function v given by (8) is t -heterogeneous. Let T be the set of atoms $\{\alpha_1, \alpha_2, \dots, \alpha_t\}$ and define the probability function v_T to be such that it gives zero probability to any $\alpha(a_j)$ for $\alpha \notin T$ whilst for

$$\bigwedge_{i=1}^p \alpha_{t_i}(a_{g_i})$$

with each $\alpha_{t_i} \in T$ and spectrum \vec{p} of length $q \leq t$,

$$\frac{t!}{(t-q)!} v_T \left(\bigwedge_{i=1}^p \alpha_{t_i}(a_{g_i}) \right) = \frac{2^n!}{(2^n - q)!} v(\vec{p}).$$

In particular when $q = t$ then,

$$\binom{2^n}{t} v(\vec{p}) = v_T \left(\bigwedge_{i=1}^p \alpha_{t_i}(a_{g_i}) \right).$$

By a similar argument to the case for 2^n -heterogeneous above, if we set

$$\rho_T(\vec{X}) = \begin{cases} |\{ \mathbf{o} \in \{1, \dots, 2^m\}^{t \times t} : |\mathcal{S}(\mathbf{o})| = t \}|^{-1} & \text{if the } X_i \subseteq T \times T \\ & \text{and the top left } t \times t \text{ submatrix of } \mathbf{x} \text{ has spectrum } \vec{1}_t, \\ 0 & \text{otherwise.} \end{cases}$$

then as above we obtain that the probability function w'_T defined by

$$w'_T \left(\bigwedge_{i,j=1}^p \beta_{r_{ij}}(a_i, a_j) \right) = \sum_{\vec{X}} \rho(\vec{X}) v_T \left(\bigwedge_{i,j=1}^p \psi_{r_{ij}}(\vec{X})(a_i, a_j) \right) \quad (23)$$

satisfies

$$w'_T(\mathbf{r}) = t! \mathcal{N}(\emptyset, \vec{1}_t)^{-1} v_T(\vec{h}) = \frac{2^n!}{(2^n - t)!} \mathcal{N}(\emptyset, \vec{1}_t)^{-1} v(\vec{h}) = d(t)^{-1} v(\vec{h})$$

for $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ with spectrum \vec{h} of length t . It follows that $w'_T = w$.

Finally to extend the result to general $\leq 2^n$ -heterogeneous w we use the representation of w as a convex sum of t -heterogeneous probability functions,

$$w = \sum_{t=1}^{2^n} \eta_t w^{[t]},$$

choose suitable $v^{[t]}$ and $\rho^{[t]}$ on disjoint languages $L^{[t]}$ for each $w^{[t]}$, take v to be the product measure of the $v^{[t]}$ on the language which is the union of the $L^{[t]}$ and finally define ρ on \vec{X} from this combined language to be

$$\sum_{t=1}^{2^n} \eta_t \rho^{[t]}$$

where $\rho^{[t]}(\vec{X}) = 0$ if the \vec{X} is not in the language $L^{[t]}$. ■

In view of our initial attempt to explain binary relations in terms of compound combinations of purely unary properties Theorem 8 is of interest because it says

that in this situation where w satisfies S_x and is $\leq t$ -heterogeneous for some t an *explanation* along these lines is always possible.

Given that we are limiting ourselves now to probability functions satisfying S_x the remaining basic case to consider is when w is homogeneous. This is the subject of our next section.

Homogeneous binary probability functions

The previous section gave a characterization of heterogeneous binary probability functions from which we can generate a host of examples. On the other hand one may feel that these are rather atypical probability functions and that the main interest lies in *homogeneous* probability functions w , that is when for all $s > 0$

$$\lim_{k \rightarrow \infty} \sum_{\vec{k} \in A_k^s} \mathcal{N}(\emptyset, \vec{k}) w(\vec{k}) = 0.$$

One obvious example of such a homogeneous probability function is the *completely independent*, or *trivial*, binary probability function which simply gives each $R_k(a_i, a_j)$ probability $1/2$ and treats them all as stochastically independent. There are also non-trivial examples as the following construction shows:

For simplicity of notation take $m = 1$, the construction being the same but with $1/2$ replaced in the obvious places by $1/2^m$ for general m . We first describe a way of generating a random $\mathbf{r} \in \{0, 1\}^{q \times q}$ and a black/red coloring of the a_1, \dots, a_q as follows. For $q = 1$ decide the color (one of black/red) for a_1 by a coin toss and similarly decide if \mathbf{r} equals $\{0\}$ or $\{1\}$. Now suppose $q > 0$ and we have constructed our random $\mathbf{r} \in \{0, 1\}^{q \times q}$ and a black/red coloring of the a_1, \dots, a_q . We describe how to extend \mathbf{r} and the coloring to $q + 1$. First decide the color of a_{q+1} by a coin toss.

If the color is decided black then (uniform) randomly pick $\mathbf{r}^+ \in \{0, 1\}^{(q+1) \times (q+1)}$ extending \mathbf{r} with the property that whenever $i < j \leq q$ and a_i, a_j both have color red then $r_{i,q+1}^+ = r_{j,q+1}^+$, $r_{q+1,i}^+ = r_{q+1,j}^+$. In other words pick \mathbf{r}^+ according to the uniform distribution from all such extensions. Notice then that in this case \mathbf{r}^+ does not kill any previous equivalence between i and j when a_i, a_j are both colored red.

If the color is decided red and no previous a_i had color red, then (uniform) randomly extend \mathbf{r} to $\mathbf{r}^+ \in \{0, 1\}^{(q+1) \times (q+1)}$, in other words giving each of the 2^{2q+1} possible extensions equal probability.

If the color is decided red and some previous a_i had color red let \mathbf{r}^+ be the unique extension of \mathbf{r} such that $I^{\mathbf{r}^+}(i, q + 1)$.

Now define $w(\mathbf{r})$ to be the probability that the above random construction generates \mathbf{r} for some coloring of a_1, \dots, a_q . We claim that w is the required binary probability function.

That w defines a binary probability function which is homogeneous but non-trivial (i.e not the completely independent solution) is clear so it is enough to check that w satisfies Sx.

To this end suppose that $\mathbf{r}, \mathbf{t} \in \{0, 1\}^{q \times q}$ have the same spectra, say their equivalence classes are in non-increasing order of size R_1, \dots, R_k , and T_1, \dots, T_k respectively, so $|R_j| = |T_j|$ for $j = 1, \dots, k$. The plan now is to show that we can pair off all possible colorings for \mathbf{r} and \mathbf{t} in such a way that both members of the pairs are equally likely to be generated as above. This is clear for the two colorings which make all of a_1, \dots, a_q black so consider a coloring associated with \mathbf{r} which makes something red. Clearly all the red a_i belong to the same I^r equivalence class, say R_j . Let it be the i_1 th, i_2 th, \dots, i_h th elements of R_j that get colored red. In this case pair this coloring with the coloring associated with \mathbf{t} which gives just the i_1 th, i_2 th, \dots, i_h th elements of T_j color red. The point now is that in both cases these two matrices with these colorings will be generated by the above construction with probability

$$\prod_{i=0}^{q-h} 2^{-2i-1},$$

which is enough to give the result.

Having demonstrated that non-trivial homogeneous binary probability functions satisfying Sx do exist we now set about deriving some of their, on occasions possibly unexpected, properties. The following result is joint with J. Landes.

Theorem 9 *If w is a homogeneous binary probability function satisfying Sx then w satisfies Regularity. More precisely if $w(\mathbf{r}) = 0$ where $|\mathcal{S}(\mathbf{r})| = t$ then w is $\leq(t-1)$ -heterogeneous.*

Proof. Let $\mathcal{S}(\mathbf{r}) = \vec{s} = \langle s_1, s_2, \dots, s_t \rangle$. By taking suitable splitting extensions of \mathbf{r} we can obtain a k such the $w(\mathbf{k}) = 0$ where $\mathcal{S}(\mathbf{k}) = \vec{1}_k$. Picking the least such k we have that w is $\leq(k-1)$ -heterogeneous. If $k \leq t$ we are done. Otherwise, by Theorem 7 we may assume that w is actually $(k-1)$ -heterogeneous (and $k-1 \geq t$). In this case let v be the unary ‘dual’ of w as in Theorem 6. Then v is $(k-1)$ -heterogeneous, $v(\vec{s}) = 0$ and $v(\vec{1}_t) > 0$.

By de Finetti’s Representation Theorem, for some measure μ on

$$\mathbb{D}_n = \left\{ \langle x_1, \dots, x_{2^n} \rangle \mid x_i \geq 0, \sum_{i=1}^{2^n} x_i = 1 \right\},$$

$$v(\vec{1}_t) = \int_{\mathbb{D}_n} x_1 x_2 \dots x_t d\mu > \epsilon > 0$$

for some ϵ , so

$$\mu\{\vec{x} \in \mathbb{D}_n \mid x_1 x_2 \dots x_t \geq \epsilon\} > 0.$$

Hence

$$\mu\{\vec{x} \in \mathbb{D}_n \mid (x_1 x_2 \dots x_t)^{s_1} \geq \epsilon^{s_1}\} > 0,$$

so

$$\mu\{\vec{x} \in \mathbb{D}_n \mid x_1^{s_1} x_2^{s_2} \dots x_k^{s_k} \geq \epsilon^{s_1}\} > 0,$$

and by de Finetti's Theorem $v(\vec{s}) > 0$, contradiction. \blacksquare

Before giving any more proofs we need some definitions.

The Splitting Property

The binary probability function w satisfies the *Splitting Property* if whenever $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ and $I^{\mathbf{r}}(1, 2)$ then

$$w(\mathbf{r}) = \lim_{q \rightarrow \infty} \sum_{\mathbf{o} \in G_q} w(\mathbf{o}),$$

where G_q is the set of $\mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q}$ such that $r_{ij} = o_{ij}$ for $1 \leq i, j \leq p$ and $\neg I^{\mathbf{o}}(1, 2)$.

The Non-splitting Property

The binary probability function w satisfies the *Non-splitting Property* if whenever $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ and $I^{\mathbf{r}}(1, 2)$ then

$$w(\mathbf{r}) > \lim_{q \rightarrow \infty} \sum_{\mathbf{o} \in G_q} w(\mathbf{o}),$$

where G_q is the set of $\mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q}$ such that $r_{ij} = o_{ij}$ for $1 \leq i, j \leq p$ and $\neg I^{\mathbf{o}}(1, 2)$.

Lemma 10 *Let w be a homogeneous binary probability function satisfying Sx and the Splitting Property. Then w is the completely independent (i.e. trivial) probability function.*

Proof. Suppose that $\sum_i p_i = \sum_i q_i = p$ and \vec{p}, \vec{q} are respectively the spectra of $\mathbf{p}, \mathbf{q} \in \{1, 2, \dots, 2^m\}^{p \times p}$. It is enough to show that $w(\vec{p}) = w(\vec{q})$.

By (6) fix large s and then large k such that

$$E_{\vec{p}} = \sum_{\vec{k} \in A_k^{\geq s}} \mathcal{N}(\vec{p}, \vec{k}) w(\vec{k})$$

is very close to $w(\vec{p})$ and

$$E_{\vec{q}} = \sum_{\vec{k} \in A_k^{\geq s}} \mathcal{N}(\vec{q}, \vec{k}) w(\vec{k})$$

is very close to $w(\vec{q})$. Now

$$E_{\vec{p}} = \sum_{\mathbf{r} \in B(\mathbf{p})} w(\mathbf{r}) + \sum_{\mathbf{r} \in C(\mathbf{p})} w(\mathbf{r})$$

where

$$\begin{aligned} B(\mathbf{p}) &= \{ \mathbf{r} \in A_k^{\geq s} : r_{ij} = p_{ij} \text{ for } i, j \leq p \text{ and } \neg I^{\mathbf{r}}(i, j) \text{ for } i < j \leq p \}, \\ C(\mathbf{p}) &= \{ \mathbf{r} \in A_k^{\geq s} : r_{ij} = p_{ij} \text{ for } i, j \leq p \text{ and } \mathbf{r} \notin B(\mathbf{p}) \}. \end{aligned}$$

Let $\mathbf{r} \in B(\mathbf{p}) \cap A_k^t$ where $s \leq t \leq k$, let us say that the smallest representatives of all the equivalence classes with respect to $I^{\mathbf{r}}$ are j_1, j_2, \dots, j_t , so in particular $j_i = i$ for $i \leq p$ since $\mathbf{r} \in B(\mathbf{p})$. Let $\hat{\mathbf{r}}$ be the matrix resulting from \mathbf{r} by sliding the j_i th row and column to the i th row and column and keeping the other rows and columns in the same order. So still $\hat{r}_{ij} = p_{ij}$ for $i, j \leq p$. Now consider a random $t \times t$ matrix \mathbf{a} with entries from $\{1, 2, \dots, 2^m\}$ agreeing with \mathbf{p} on its top left $p \times p$ submatrix. The proportion of such choices of \mathbf{a} such that $I^{\mathbf{a}}(i, j)$ holds for particular $i \neq j$ is less than $2^{-2m(t-p)}$, so the proportion such that $I^{\mathbf{a}}(i, j)$ holds for *some* $i \neq j$ is less than $t^2 2^{-2m(t-p)}$, which in turn is uniformly bounded by $s^2 2^{-2m(s-p)}$ ($= \epsilon$ say). Hence

$$|\{ \hat{\mathbf{r}} : \mathbf{r} \in B(\mathbf{p}) \cap A_k^t \}| \quad \text{and} \quad |\{ \hat{\mathbf{r}} : \mathbf{r} \in B(\mathbf{q}) \cap A_k^t \}|$$

are the same to within this small fraction ϵ of their sum. So we can split these sets into subsets $D_1^{\vec{p}}, D_2^{\vec{p}}$ and $D_1^{\vec{q}}, D_2^{\vec{q}}$ such that $|D_1^{\vec{p}}| = |D_1^{\vec{q}}|$ and $|D_2^{\vec{p}}| \leq \epsilon |D_1^{\vec{p}}|$, $|D_2^{\vec{q}}| \leq \epsilon |D_1^{\vec{q}}|$. Clearly

$$|\{ \mathbf{r} \in B(\mathbf{p}) \cap A_k^t : \hat{\mathbf{r}} \in D_1^{\vec{p}} \}| = |\{ \mathbf{r} \in B(\mathbf{q}) \cap A_k^t : \hat{\mathbf{r}} \in D_1^{\vec{q}} \}| \quad (24)$$

so by Sx the sums of the $w(\mathbf{r})$ for \mathbf{r} in these sets will be the same. As for the $\hat{\mathbf{r}} \in D_2^{\vec{p}}$, summing the corresponding $w(\mathbf{r})$ will yield a sum which is at most ϵ times that obtained by summing the $w(\mathbf{r})$ for $\hat{\mathbf{r}} \in D_1^{\vec{p}}$. Finally summing over the $s \leq t \leq k$ shows that

$$\sum_{\mathbf{r} \in B(\mathbf{p})} w(\mathbf{r}) \quad \text{and} \quad \sum_{\mathbf{r} \in B(\mathbf{q})} w(\mathbf{r})$$

are very close.

The proof now reduces to showing that

$$\sum_{\mathbf{r} \in C(\mathbf{p})} w(\mathbf{r}) \quad \text{and} \quad \sum_{\mathbf{r} \in C(\mathbf{q})} w(\mathbf{r})$$

are also very close. But this is true by the assumption on w . ■

Corollary 11 *Let w be a non-trivial homogeneous binary probability function satisfying Sx. Then w satisfies the Non-splitting Property.*

Proof. Since w is non-trivial, by Lemma 10, for *some* $\mathbf{r}' \in \{1, 2, \dots, 2^m\}^{p' \times p'}$ with $I_0^{\mathbf{r}'}(1, 2)$,

$$w(\mathbf{r}') > \lim_{q \rightarrow \infty} \sum_{\mathbf{o} \in G'_q} w(\mathbf{o}),$$

where G'_q is the set of $\mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q}$ such that $r'_{ij} = o_{ij}$ for $1 \leq i, j \leq p'$ and $\neg I^{\mathbf{o}}(1, 2)$. Clearly we may take p' as large as we wish here. Now let $\mathbf{r} \in \{1, 2, \dots, 2^m\}^{p \times p}$ and $I^{\mathbf{r}}(1, 2)$. Then by taking p' sufficiently large compared with p we see that there will be some extension $\mathbf{r}'' \in \{1, 2, \dots, 2^m\}^{p' \times p'}$ of \mathbf{r} such that $I^{\mathbf{r}''}(1, 2)$, $\mathcal{S}(\mathbf{r}') = \mathcal{S}(\mathbf{r}'')$ and the equivalence classes of 1 with respect to both $I^{\mathbf{r}'}$ and $I^{\mathbf{r}''}$ have the same size. Then by Sx

$$w(\mathbf{r}'') > \lim_{q \rightarrow \infty} \sum_{\mathbf{o} \in G''_q} w(\mathbf{o}),$$

so clearly this same strict inequality must hold for \mathbf{r} . ■

For future reference notice that 1 and 2 in this Corollary can be replaced by any other distinct $i, j \leq p$.

Corollary 11 seems to us surprising because it says in effect that if any two individuals look different in almost all possible futures according to a homogeneous w satisfying Sx then w must be the completely independent solution. In other words, if w is to avoid being the completely independent solution it must allow the real possibility that individuals that look identical today will remain looking identical for all time.

The non-trivial homogeneous probability function constructed above of course satisfies Non-splitting. It also satisfies the following property:

The Clustering Property

The binary probability function w satisfies the *Clustering Property* if whenever $\mathbf{t} \in \{1, 2, \dots, 2^m\}^{p \times p}$ then there is some $\epsilon > 0$ such that

$$\lim_{q \rightarrow \infty} \sum_{\mathbf{o} \in H_q} w(\mathbf{o}) > 0$$

where H_q is the set of $\mathbf{o} \in \{1, 2, \dots, 2^m\}^{q \times q}$ such that $t_{ij} = o_{ij}$ for $1 \leq i, j \leq p$ and $|\{j \leq q : I^{\mathbf{o}}(1, j)\}| \geq \epsilon q$.

Loosely speaking then, this is saying that there is a non-zero probability that in the future at least one in any N of the individuals observed will be identical to the the first (or any other fixed) individual observed.

As we now show this is no accident, it must hold for any non-trivial homogeneous w satisfying Sx.

Theorem 12 *Let w be a non-trivial homogeneous probability function satisfying Sx. Then w satisfies the Clustering Property.*

Proof. We shall first show this result for \mathbf{t} empty (i.e. $p = 0$), the result for general \mathbf{t} following by restricting all matrices to extensions of \mathbf{t} .

By non-splitting it follows that there is a $\delta > 0$ such that for all $q \geq 2$

$$\sum_{\substack{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q} \\ \neg I^{\mathbf{r}}(1,2)}} w(\mathbf{r}) \geq \delta. \quad (25)$$

Clearly by Ex we could replace 1,2 in this inequality by any $i < j \leq q$.

Now consider the sum

$$\sum_{i < j \leq q} \sum_{\substack{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q} \\ \neg I^{\mathbf{r}}(i,j)}} w(\mathbf{r}).$$

For $\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q}$ with spectrum $\langle s_1, s_2, \dots, s_k \rangle$ the term $w(\mathbf{r})$ appears in this sum once for every $i < j \leq q$ such that $\neg I^{\mathbf{r}}(i, j)$, in other words $\sum_{g < h \leq k} s_g s_h (= \iota(\mathbf{r})$, say) times. Hence

$$\sum_{i < j \leq q} \sum_{\substack{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q} \\ \neg I^{\mathbf{r}}(i,j)}} w(\mathbf{r}) = \sum_{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q}} \iota(\mathbf{r}) w(\mathbf{r}). \quad (26)$$

By (25) and Sx the left hand side of this equation is at most $q(q-1)(1-\delta)/2$ so

$$q(q-1)(1-\delta)/2 \geq \sum_{\mathbf{r} \in \{0,1\}^{q \times q}} \iota(\mathbf{r}) w(\mathbf{r}). \quad (27)$$

We now claim that for $\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q}$ with $\mathcal{S}(\mathbf{r}) = \langle s_1, s_2, \dots, s_k \rangle$,

$$\iota(\mathbf{r}) \geq q(q-s_1)/2. \quad (28)$$

To see this notice that $\sum_{i=1}^k s_i^2$ is bounded by the maximum of the real function $\sum_{i=1}^q x_i^2$ subject to the constraints $0 \leq x_i \leq s_1$, $\sum_{i=1}^q x_i = q$. This maximum is taken (in particular) when $x_i = s_1$ for $i = 1, \dots, d$, $x_{d+1} = q - ds_1$ and $x_i = 0$ otherwise, where $d = \lceil (q/s_1) \rceil$. Hence

$$\sum_{i=1}^k s_i^2 \leq ds_1^2 + (q - ds_1)^2$$

and

$$\begin{aligned}
2\iota(\mathbf{r}) &= \sum_{i \neq j} s_i s_j \\
&= q^2 - \sum_i s_i^2 \\
&\geq q^2 - (ds_1^2 + (q - ds_1)^2) \\
&= ds_1(2q - (d+1)s_1) \\
&= (q - u)(q - (s_1 - u)) \quad \text{where } 0 \leq u = q - ds_1 \leq s_1 \\
&\geq q(q - s_1).
\end{aligned}$$

Now suppose that for any $\epsilon > 0$ we could find arbitrarily large q such that

$$\sum_{\substack{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q} \\ s_1/q \geq \epsilon}} w(\mathbf{r}) < \epsilon.$$

Then

$$\begin{aligned}
\sum_{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q}} \iota(\mathbf{r})w(\mathbf{r}) &\geq \sum_{\substack{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q} \\ s_1/q < \epsilon}} \iota(\mathbf{r})w(\mathbf{r}) \\
&\geq (1 - \epsilon)q(q - \epsilon q)/2 \\
&= q^2(1 - \epsilon)^2/2.
\end{aligned}$$

But taking ϵ sufficiently small and q sufficiently large this contradicts (27). It follows then that for some $\epsilon > 0$ and all q eventually

$$\sum_{\substack{\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q} \\ s_1/q \geq \epsilon}} w(\mathbf{r}) \geq \epsilon.$$

But by Ex, amongst the $\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q}$ with $s_1/q \geq \epsilon$ the probability that $1 \in s_1$ is at least ϵ so overall the probability that $|\{j \leq q : I^{\mathbf{r}}(1, j)\}| \geq \epsilon q$ for $\mathbf{r} \in \{1, \dots, 2^m\}^{q \times q}$ is at least ϵ^2 , as required. ■

Corollary 13 *If w is a non-trivial binary probability function satisfying Sx then w satisfies Non-splitting and Clustering.*

Proof. These clearly hold for w t -heterogeneous and, by Corollary 11 and Theorem 12, for non-trivial homogeneous w also. The result now follows from the Representation Theorem 7. ■

Conclusions

In this paper we have made some tentative first steps towards understanding induction for binary relations, a problem which has (apparently) been strangely neglected following the successes of half a century ago of Carnap, Johnson et al (see [1],[2],[19]) with unary relations (i.e.properties). For the most part we have investigated probability functions on finite binary languages satisfying Spectrum Exchangeability, Sx , a property which naturally generalizes Atom Exchangeability for unary languages (and in turn itself generalizes in a natural way to higher arities still, see the forthcoming [25]). We have shown that probability functions satisfying Sx have some pleasing properties (Vx , UC , and Non-splitting and Clustering in the non-trivial case) and have derived some representation results.

Many avenues for future investigations remain it seems largely unexplored. Are there alternative principles to Sx which also have intuitive appeal in this context? Even assuming Sx there are still numerous open questions, in particular the fundamental question of ‘instantial relevance’, how does, or should, the past impinge on the future? - a question which surely underlies the whole study of inductive logic(s).

In the unary case de Finetti’s Representation Theorem provides a straightforward route to proving a number of ‘instantial relevance’ results and one might look to such a result here to provide a similar service. In [24] Krauss proves a representation theorem for exchangeable binary probability functions (i.e. satisfying Ex) in the case of a language with a single binary predicate although the corresponding corollary for instantial relevance that Krauss obtains requires the pairs of constants witnessing previous instances to be disjoint and is therefore apparently not applicable to cases such as the one above. Indeed because the witnessing pairs are to be disjoint the direct version of instantial relevance given by Krauss’s representation can already be obtained directly by applying de Finetti’s theorem to a suitable, essentially unary, fragment of the language. Subsequently Hoover [17] (see also [20], [21]) provided a possibly more user friendly representation theorem for this class of probability functions, though it currently appears to be most useful as a source of counter examples to the sort of instantial relevance principles one might initially have expected. Hopefully the recent formulation of a de Finetti theorem¹⁴ for heterogeneous binary probability functions satisfying Sx will help clarify the situation as regards such principles in that case.

A further generalization is to infinite languages, allowing infinitely many binary predicates in the language. In this paper we have limited our attention to finite languages because this is very much a pilot study where we would wish to keep matters as simple as possible. However we know from the unary situation

¹⁴Suggested by methods of Matúš on Bernoulli Shifts [28] and to be presented in a forthcoming paper.

that restricting attention to finitely many predicates can result in rather different intuitions (and theorems), whilst of the two the assumption that there are (potentially at least) infinitely many predicates seems closer to the real world from which we draw our intuitions than the assumption of a fixed finite bound thereon.

Acknowledgements

We would like to thank the referee (unknown) and editor, Horacio Costa, for their patience and very useful comments, František Matúš for alerting us to relevant literature in Probability Theory and finally Jürgen Landes and Alena Vencovská for suggesting a number of improvements to the original version.

References

- [1] Carnap, R., *Logical Foundations of Probability*, University of Chicago Press, Chicago, Routledge & Kegan Paul Ltd., 1950.
- [2] Carnap, R., *The Continuum of Inductive Methods*, University of Chicago Press, 1952.
- [3] Carnap, R., A basic system of inductive logic, in *Studies in Inductive Logic and Probability*, Volume II, ed. R. C. Jeffrey, University of California Press, 1980, 7-155.
- [4] Carnap, R., On the application of inductive logic, *Philosophy and Phenomenology Research*, 1947, **8**:133-147.
- [5] Carnap, R., Reply to Nelson Goodman, *Philosophy and Phenomenology Research*, 1947, **8**:461-462.
- [6] Coletti, G. & Scozzafava, R., *Probabilistic Logic in a Coherent Setting*, Trends in Logic, **15**, Kluwer Academic Press, London, Dordrecht, 2002.
- [7] de Finetti, B., On the condition of partial exchangeability, *Studies in Inductive Logic and Probability*, Volume II, Ed. R.C. Jeffrey, University of California Press, Berkley and Los Angeles, 1980.
- [8] de Finetti, B., *Theory of Probability*, Vol. 1, Wiley, New York, 1974.
- [9] Fitelson, B., *Inductive Logic*, <http://fitelson.org/il.pdf>
- [10] Gaifman, H., Concerning measures on first order calculi, *Israel Journal of Mathematics*, 1964, **2**:1-18.

- [11] Gaifman, H. & Snir, M., Probabilities over rich languages, *Journal of Symbolic Logic*, 1982, **47**(3):495-548.
- [12] Glaister, S., Inductive Logic, in *A Companion to Philosophical Logic*, ed. D. Jacquette, Blackwell, London, 2001, 565-581.
- [13] Goodman, N., *Fact, Fiction and Forecast*, Cambridge MA: Harvard UP, 4th edition, 1983.
- [14] Goodman, N., A query on confirmation, *Journal of Philosophy*, 1946, **43**:383-385.
- [15] Goodman, N., On infirmities in confirmation-theory, *Philosophy and Phenomenology Research*, 1947, **8**:149-151.
- [16] Hill, M.J., Paris, J.B. & Wilmers, G.M., Some observations on induction in predicate probabilistic reasoning, *Journal of Philosophical Logic*, 2002, **31**(1):43-75.
- [17] Hoover, D.N., Relations on probability spaces and arrays of random variables. Preprint, Institute of Advanced Study, Princeton, 1979.
- [18] Jeffrey, R.C., Goodman's query, *Journal of Philosophy*, 1966, **63**(11):281-288.
- [19] Johnson, W.E., Probability: The deductive and inductive problems, *Mind*, 1932, **41**(164):409-423.
- [20] Kallenberg, O., *Probabilistic Symmetries and Invariance Principles*, Springer, USA, ISBN-10: 0-387-25115-4, 2005.
- [21] Kallenberg, O., The Ottawa Workshop,
<http://www.mathstat.uottawa.ca/~givanoff/wskallenberg.pdf>
- [22] Kawalec, P., Back to green as perspectives on confirmation, *Justification, Truth and Belief*, <http://www.jtb-forum.pl>, January, 2002.
- [23] Kemeny, J.G., Carnap's theory of probability and induction, in *The Philosophy of Rudolf Carnap*, ed. P.A.Schilpp, La Salle, Illinois, Open Court, 1963, 711-738.
- [24] Krauss, P.H., Representation of symmetric probability models, *Journal of Symbolic Logic*, 1969, **34**(2):183-193.
- [25] Landes, J., Doctorial Thesis, Manchester University, UK, to appear.
- [26] Maher, P., Probabilities for two properties, *Erkenntnis*, 2000, **52**:63-91.

- [27] Maher, P., Probabilities for multiple properties: The models of Hesse, Carnap and Kemeny, *Erkenntnis*, 2001, **55**:183-216.
- [28] Matúš, F., Block-factor fields of Bernoulli shifts, *Proceedings of Prague Stochastics'98*, Vol.II, 1998, 383-389.
- [29] Miller, D., Popper's qualitative theory of verisimilitude, *British Journal for the Philosophy of Science*, 1974, **25**:166-177.
- [30] Nix, C.J., *Probabilistic Induction in the Predicate Calculus* Doctorial Thesis, Manchester University, Manchester, UK, 2005. See <http://www.maths.man.ac.uk/~jeff/#students>
- [31] Nix, C.J. & Paris, J.B., 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.
- [32] Paris, J.B., *The Uncertain Reasoner's Companion*, Cambridge University Press, 1994.
- [33] Russell, B., *Principles of Mathematics*, George Allen & Unwin Ltd, London, 2nd edition, 1937.
- [34] Scott, D. & Krauss, P., Assigning probabilities to logical formulas, in *Aspects of Inductive Logic*, eds. J.Hintikka & P.Suppes, North-Holland, Amsterdam, 1966, 219-264.
- [35] Stalker, D. ed., *Grue! The New Riddle of Induction*, Open Court, 1994.