

The Counterpart Principle of Analogical Support by Structural Similarity

A.Hill* and J.B.Paris
School of Mathematics
The University of Manchester
Manchester M13 9PL

alex.hill@gmail.com
jeff.paris@manchester.ac.uk

August 21, 2013

Abstract

We propose and investigate an *Analogy Principle* in the context of Unary Inductive Logic based on a notion of support by structural similarity which is often employed to motivate scientific conjectures.

Key words: Analogy, Inductive Logic, Probability Logic, Uncertain Reasoning.

Introduction

Starting with the founding work of Carnap, [1], there have been a number of attempts within Inductive Logic to formally capture the idea of ‘analogical support’ between evidence and hypotheses, for example [5], [16], [27], [28]. These attempts to date do not seem to us to have provided a comprehensive formal justification to explain the intuitive appeal of ‘support by analogy’, and indeed our earlier paper

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

[10] shows that some of these formalizations have very limited applicability. By contrast in this paper we shall present an analogy principle within the context of Inductive Logic which holds widely, in particular it holds for the continua of inductive methods of Carnap and Nix-Paris. This principle is inspired by an alternative account of analogical support based on structural similarity.

The structure of this paper is that after a section introducing the formal context and notation. we shall briefly motivate the notion of analogy, structural similarity, which we intend to formalize and investigate. Such a formalization, the *Counterpart Principle* as we shall call it, together with an investigation into conditions under which it does/does not hold will then follow.

Context and Notation

We will be working, as usual, in the first order framework for (unary) Inductive Logic¹ where we have a predicate language L with finitely many, say q , (unary) predicate symbols P_1, \dots, P_q , constant symbols a_1, a_2, a_3, \dots , the intention being that these constants enumerate the universe, and no function symbols nor equality. Let SL denote the set of sentences of L and $QFSL$ the quantifier free sentences of L . We will use θ, ϕ, ψ etc. for elements of SL and adopt throughout the convention that if we write a sentence as $\theta(a_{i_1}, a_{i_2}, \dots, a_{i_n})$ then all the constants appearing in θ are amongst these $a_{i_1}, a_{i_2}, \dots, a_{i_n}$, though they need not all actually appear.

In this context, where the a_i are intended to enumerate the universe, we define a *Probability Function on L* to be a map $w : SL \rightarrow [0, 1]$ such that for all $\theta, \phi, \exists x \psi(x) \in SL$:

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

¹Our general approach to Inductive Logic, see for example [21], [23], is actually close to what Carnap in [3] termed *Pure Inductive Logic*, the intention of ‘Pure’ here being similar to that in ‘Pure Mathematics’. That is, our aim is to investigate the logic itself in isolation, devoid of any particular intended real world interpretation. Of course the rational principles with which Pure Inductive Logic is concerned are (and indeed should be if the subject is not to become simple mathematics for its own sake) almost invariably *motivated* by real world examples. But once such principles are formulated it is then the task of Pure Inductive Logic to investigate their consequences as they stand, without borrowing further from any particular, special, interpretation. It is akin to studying a differential equation which has arisen in modeling, say, a bridge. Its solutions apply to any model for which the equation is appropriate, and if they do not apply to the particular bridge we started off with then that is because the original formulation was somehow lacking, not that the differential equation itself is at fault.

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

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

The following theorem, due to Gaifman, [8], shows that a probability function w is actually already determined by its action on *QFSL*.

Theorem 1. *Let $w : QFSL \rightarrow [0, 1]$ satisfy (P1), (P2) for $\theta, \phi \in QFSL$. Then w has a unique extension to *SL* satisfying (P1-3).*

The aim in Inductive Logic, as we view it, is to pick out probability functions on L which are arguably *logical* or *rational* in the sense that they could be the choice of a rational agent. Or to put it another way to discard probability functions which could be judged in some sense to be ‘irrational’.² The usual method of thinning down towards such rational choices is to impose ‘rationality principles’ which these probability functions should arguably satisfy. Of course there can be considerable disagreement about which principles are rational (to the extent of different candidates being mutually inconsistent, see for example [20]) but one such widely accepted principle is that the inherent symmetry between the constants should be respected by any rational probability function w on L . Precisely w should satisfy:

The Constant Exchangeability Principle (Ex)

For $\theta, \theta' \in QFSL$, if θ' is obtained from θ by replacing the constant symbols³ $a_{i_1}, a_{i_2}, \dots, a_{i_n}$ in θ by $a_{k_1}, a_{k_2}, \dots, a_{k_n}$, respectively, then $w(\theta) = w(\theta')$.

It is worth remarking that Ex in fact implies this same principle even for $\theta, \theta' \in SL$. Thus in Theorem 1 if w satisfies Ex then so will its extension to *SL*.

The *atoms*⁴ of the language L are the formulae $\alpha_i(x)$, $i = 1, 2, \dots, 2^q$, of the form

$$\pm P_1(x) \wedge \pm P_2(x) \wedge \dots \wedge \pm P_q(x)$$

where $\pm P_j$ stands for either P_j or $\neg P_j$.

The atoms are disjoint and exhaustive, so if w is a probability function on L then by (P1-2), for any m ,

$$w(\alpha_1(a_m)), w(\alpha_2(a_m)), \dots, w(\alpha_{2^q}(a_m)) \geq 0 \quad \text{and} \quad \sum_i w(\alpha_i(a_m)) = 1.$$

²It is interesting that ‘irrationality’ seems much easier to spot than ‘rationality’.

³In such lists all the terms will be assumed to be distinct unless otherwise stated.

⁴Corresponding to Carnap’s Q -predicates within the present framework.

Conversely for any such vector of non-negative numbers x_i with sum 1 there is a probability function w on L with $w(\alpha_i(a_m)) = x_i$ for $i = 1, 2, \dots, 2^q$. Namely for

$$\vec{x} \in \mathbb{D}_{2^q} = \{\langle x_1, \dots, x_{2^q} \rangle \mid x_i \geq 0, \sum_{i=1}^{2^q} x_i = 1\}$$

define $w_{\vec{x}}$ on state descriptions, that is sentences of the form $\bigwedge_{i=1}^n \alpha_{h_i}(a_{k_i})$, by

$$w_{\vec{x}} \left(\bigwedge_{i=1}^n \alpha_{h_i}(a_{k_i}) \right) = \prod_{i=1}^n x_{h_i} = \prod_{j=1}^{2^q} x_j^{n_j}$$

where n_j is the number of times that the atom α_j occurs amongst the α_{h_i} . Using the Disjunctive Normal Form Theorem it is easy to see that $w_{\vec{x}}$ can be extended to $QFSL$ to satisfy (P1-2) and Ex and hence by Theorem 1 to a probability function on L , which continues to satisfy Ex. Clearly also $w_{\vec{x}}(\alpha_i(a_m)) = x_i$ for $i = 1, 2, \dots, 2^q$ (and any a_m).

In what follows we take Ex as a standing assumption. This allows us access to a powerful representation theorem due to de Finetti, see [7].⁵

De Finetti's Representation Theorem

If the probability function w on L satisfies Ex then there is a (countably additive) measure μ on \mathbb{D}_{2^q} such that for $\theta \in SL$,

$$w(\theta) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta) d\mu(\vec{x}). \quad (1)$$

Conversely if w is defined by (1) then w is a probability function on L satisfying Ex.

We refer to the measure μ here as the *de Finetti prior* of w .

In [9] Gaifman showed the following consequence of de Finetti's Representation Theorem:⁶

⁵As given here the converse direction assumes a result due to Gaifman on restrictions of probability functions, see [8].

⁶One may have hoped that such an elementary and comprehensible statement as this theorem would have a correspondingly elementary and comprehensible proof, using only simple applications of (P1-2) say. Unfortunately we know of no such proof and quite commonly it appears that in this subject we cannot avoid diversions into 'higher mathematics'. We will later see an (apparently) similar situation with the use of Theorems 10, 11 to conclude on page 17 that Carnap's c_λ , for $0 < \lambda < \infty$, satisfy our forthcoming Counterpart Principle.

Theorem 2. *If the probability function w on L satisfies Ex then it satisfies:*

The Principle of Instantial Relevance (PIR)

For $\psi \in QFSL$, α an atom of L and a_i, a_j distinct constant symbols not mentioned in ψ ,⁷

$$w(\alpha(a_i) \mid \alpha(a_j) \wedge \psi) \geq w(\alpha(a_i) \mid \psi). \quad (2)$$

In short the additional evidence $\alpha(a_j)$ enhances (or at least does not decrease) the probability of $\alpha(a_i)$ conditional on evidence ψ .

Extending the idea of symmetry between symbols of the language, we might also feel it rational to require that the predicates be exchangeable.

The Predicate Exchangeability Principle (Px)

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

Using Theorem 1 it is easy to show that Px implies that the same property holds even when θ, θ' are sentences of L rather than just quantifier free sentences, a fact that we shall use without further mention in what follows.

The following is a stronger symmetry based principle which is also frequently seen in Inductive Logic; for example, it is satisfied by both Carnap's Continuum of Inductive Methods and the Nix-Paris Continuum, [20].

The Atom Exchangeability Principle (Ax)

If σ is a permutation of $1, 2, \dots, 2^q$, then

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

Note that Ax (with Ex) implies that the probability of a *state description* $\theta = \bigwedge_{i=1}^n \alpha_{h_i}(a_{j_i})$ depends only on the multiset⁸ $\{n_1, n_2, \dots, n_{2^q}\}$ where n_j is the number of times that the atom α_j appears amongst the α_{h_i} . We call this multiset the *spectrum* of θ . It follows then that stated in this form Ax implies Px.

⁷In order to circumvent any problems when the conditioning sentence has probability zero we take an expression such as $w(\phi_1 \mid \psi_1) \geq w(\phi_2 \mid \psi_2)$ to be short for the inequality formed by multiplying out the denominators, i.e. $w(\phi_1 \wedge \psi_1) \cdot w(\psi_2) \geq w(\phi_2 \wedge \psi_2) \cdot w(\psi_1)$ in this case. So PIR as defined will automatically hold when $w(\alpha(a_j) \wedge \psi) = 0$ since both sides of the inequality will then be zero.

⁸A multiset is just a set in which we allow the same element to appear more than once. The usual convention here is that n_j which are zero are omitted from mention in the spectrum. However in this case it will be convenient to include them.

In addition to symmetry there have been numerous attempts to incorporate principles based on analogy into Inductive Logic, initially by Carnap, for example in [4], by Carnap & Stegmüller [5] and later for example by Festa [6], Kuipers [13], Maher [16], [17], Maio [18], Niiniluoto [19], Romeijn [27], Skyrms [28]. Generally these have considered analogy as deriving from the sharing of similar or identical properties by the constants. In other words, a principle has been sought based on a notion of *distance* between atoms. This approach means that Ax is violated. Although this may not be a failing of such principles it is interesting to note that the alternative conception of analogy presented in this paper is consistent with Ax and is widely satisfied, including by the probability functions of the afore-mentioned continua.

Support by Structural Similarity

As a toy example⁹ of what we have in mind by ‘support by structural similarity’ suppose that I am observing windfall pears and the first three I see are all green. Then that should not decrease my belief that the next three pears observed will all have a maggot, since my first observation of the three green pears suggests that these pears are a pretty uniform bunch, so no less likely to *all* have a maggot than was the case before I had observed the three green pears.

For an everyday example consider the assertion that there is, or at least once has been, life on Mars. Most of us (we imagine) would think this was indeed worth the multi-billion dollars so far spent trying to confirm it. Equally however for most of us this is (we again imagine) largely based on the similarity between the Earth and Mars. They are both planets with fairly similar ages and orbits, and both have atmospheres of sorts. On top of that ‘Earth-Life’ is abundant almost across the globe so isn’t ‘Mars-Life’ just a missing piece in the jigsaw, in the pattern, something almost to be expected?

Interestingly we can give here a third example which directly relates to Theorem 3 of this paper. As we have seen in the presence of Constant Exchangeability, Ex, we have the Principle of Instantial Relevance, PIR, saying that, in the notation of (2), the additional evidence $\alpha(a_j)$ provides support for $\alpha(a_i)$. In that case we might conjecture that there should be some similar ‘Support Principle’ when we have Predicate Exchangeability in place of Constant Exchangeability. As we shall see this is exactly the case.

This last example is by no means uncommon within our experience of doing mathematics. In fact with most conjectures one makes there is somewhere in the back-

⁹Arising out of the discussion with one of the reviewers.

ground a structurally similar situation in which the corresponding conjecture has been confirmed. This is not to say that this is invariably the right goal to aim for, quite often one's conjecture based on this form of analogy turns out to be false, but what it does seem to show is that in forming and attempting to confirm such conjectures we are implicitly giving credence to a principle of support by structural similarity. Nor, of course, do our experiences here seem to differ from those of our fellow mathematicians, or even scientists in general – in [25], [26] Polya gives a nice account of just such analogical reasoning.

The Counterpart Principle

In view of the above discussion we are lead to propose, within the context of PIL, the following as a principle of analogical reasoning:

Counterpart Principle (CP)¹⁰

For any $\theta \in SL$, if $\theta' \in SL$ is obtained by replacing the predicate and constant symbols appearing in θ by (distinct) new ones not occurring in θ then

$$w(\theta | \theta') \geq w(\theta). \quad (3)$$

Our plan now in this section is to show that the Counterpart Principle, CP, is rather widely satisfied. In the section which follows we will make some observations on when we can have strict inequality in (3).

¹⁰Superficial this principle might be thought to resemble the Analogieschluss ('inference by induction') of Carnap & Stegmüller dating back to their [5, p226]. The intuition here is that for state descriptions

$$\psi(a_1, \dots, a_n) = \bigwedge_{i=1}^n \alpha_{h_i}^{L_1}(a_i), \quad \phi(a_1, \dots, a_n) = \bigwedge_{i=1}^n \alpha_{g_i}^{L_2}(a_i)$$

for disjoint languages L_1, L_2 the probability of the state description $\psi(a_1, \dots, a_n) \wedge \phi(a_1, \dots, a_n)$ of $L_1 \cup L_2$ should be greater the more often

$$h_i = h_j \iff g_i = g_j.$$

The difference here with CP is that here the analogy or link is between the relations of indistinguishability between the same a_i engendered by ψ and ϕ whereas in CP it is between the forms of sentences θ, θ' involving disjoint a_i . (For more on Carnap & Stegmüller's Analogieschluss within the approach to Inductive Logic taken here see [15], [23].)

We need the following notion: A probability function w on a language L is said to satisfy *Unary Language Invariance*, ULi, if there is a family of probability functions $w^{\mathcal{L}}$, one on each unary language \mathcal{L} , each satisfying Px, such that $w^L = w$ and whenever $\mathcal{L} \subset \mathcal{L}'$ then $w^{\mathcal{L}'}$ restricted to $S\mathcal{L}$ (notice that $S\mathcal{L} \subseteq S\mathcal{L}'$) equals $w^{\mathcal{L}}$. We say that w satisfies ULi *with* Ax if in addition we can choose these $w^{\mathcal{L}}$ to satisfy Ax.

Our argument in favour of it being rational for a probability function w on L to satisfy ULi is that it appears reasonable to suppose that were we to enlarge the language then w could be extended to this larger language, after all it would seem unjustified to assume from the start that L was all the language there could ever be (a point made by Kemeny already in [12]). Similarly if we are assuming the position that \mathcal{P} is a property our rational choice of probability function w on L should possess then we should equally demand this property of our chosen extension to larger languages. This then provides a rational justification for Language Invariance. In particular we should require that the probability functions in this family satisfy Px (and Ex by standing assumption), thus yielding ULi.

Note that ULi equivalently means that w can be extended to a probability function w_∞ on the infinite language $L_\infty = \{P_1, P_2, P_3, \dots\}$ satisfying Px. In more detail, given the ULi family $\{w^{\mathcal{L}}\}$ we let L_n be the language with predicate symbols $\{P_1, P_2, \dots, P_n\}$ and for $\theta \in SL_\infty$ set $w_\infty(\theta) = w^{L_n}(\theta)$ for any n sufficiently large that $\theta \in SL_n$. Conversely, given such a probability function w_∞ and a language L with predicate symbols $\{R_1, R_2, \dots, R_n\}$ we can define the required $w^L(\theta)$ to be $w_\infty(\theta')$ where θ' is the result of replacing each R_i in θ by P_i . (Notice that by Px this is independent of how we list the predicate symbols of L .)

Theorem 3. *Let w satisfy ULi. Then w satisfies the Counterpart Principle, CP.*

Proof. Assume that w satisfies ULi and let w_∞ be a probability function on the infinite (unary) language $L_\infty = \{P_1, P_2, P_3, \dots\}$ extending w and satisfying Px. Let θ, θ' be as in the statement of CP, without loss of generality assume that all the constant symbols appearing in θ are amongst a_1, a_2, \dots, a_k , all the relation symbols appearing in θ are amongst P_1, P_2, \dots, P_j , and for θ' they are correspondingly

$$a_{k+1}, a_{k+2}, \dots, a_{2k}, \quad P_{j+1}, P_{j+2}, \dots, P_{2j}.$$

So we can write

$$\begin{aligned} \theta &= \theta(a_1, a_2, \dots, a_k, P_1, P_2, \dots, P_j), \\ \theta' &= \theta(a_{k+1}, a_{k+2}, \dots, a_{2k}, P_{j+1}, P_{j+2}, \dots, P_{2j}). \end{aligned}$$

With this in place let

$$\theta_{i+1} = \theta(a_{ik+1}, a_{ik+2}, \dots, a_{(i+1)k}, P_{ij+1}, P_{ij+2}, \dots, P_{(i+1)j}) \in SL_\infty$$

so $\theta_1 = \theta$, $\theta_2 = \theta'$. Let \mathcal{L} be the unary language with a single unary relation symbol P and define $\tau : QFSL \rightarrow SL_\infty$ by

$$\tau(P(a_i)) = \theta_i, \quad \tau(\neg\phi) = \neg\tau(\phi), \quad \tau(\phi \wedge \psi) = \tau(\phi) \wedge \tau(\psi), \quad \text{etc.}$$

for $\phi, \psi \in QFSL$.

Now define $v : QFSL \rightarrow [0, 1]$ by

$$v(\phi) = w_\infty(\tau(\phi)).$$

Then since w_∞ satisfies (P1-2) (on SL_∞) so does v (on $QFSL$). Also since w_∞ satisfies Ex + Px, for $\phi \in QFSL$, permuting the θ_i in $w(\tau(\phi))$ will leave this value unchanged so permuting the a_i in ϕ will leave $v(\phi)$ unchanged. i.e. v satisfies Ex.

By Theorem 1 v has an extension to a probability function on SL satisfying Ex and hence satisfying PIR by Theorem 2. In particular then

$$v(P(a_1) | P(a_2)) \geq v(P(a_1)).$$

But since $\tau(P(a_1)) = \theta$, $\tau(P(a_2)) = \theta'$ this amounts to

$$w_\infty(\theta | \theta') \geq w_\infty(\theta)$$

and hence gives the Counterpart Principle for w since w_∞ agrees with w on SL . □

We remark that by a small amendment of the proof this conclusion, CP, can be strengthened to require that only *some* of the constant and predicate symbols in θ are changed when forming θ' . Furthermore by then using w_∞ conditioned on ψ in place of w_∞ through this proof we can further strengthen the conclusion to

$$w(\theta | \theta' \wedge \psi) \geq w(\theta | \psi)$$

for any $\psi \in SL$ mentioning only constant and predicate symbols common to both θ and θ'

The condition ULi required for Theorem 3 holds for the c_λ^L of Carnap's Continuum of Inductive Methods, [2], and also for the w_L^δ of the Nix-Paris Continuum¹¹ defined

¹¹See [20] for an explanation of how this continuum, like Carnap's Continuum of Inductive Methods, is singled out by arguably rational principles.

by

$$w_L^\delta = 2^{-q} \sum_{j=1}^{2^q} w_{\vec{e}_j}$$

where $\vec{e}_j = \langle \gamma, \gamma, \dots, \gamma, \gamma + \delta, \gamma, \dots, \gamma, \gamma \rangle$, the δ occurring in the j th coordinate, $\gamma = 2^{-q}(1 - \delta)$ and $0 \leq \delta \leq 1$. Indeed they both satisfy the stronger condition of ULi with Ax, the language invariant families being given by fixing λ , respectively δ , and varying L , see [20], [21], or for a more technical account [14] or [23].

It is worth noting that we cannot do without ULi here, Ex and Px alone do not guarantee that a probability function satisfies CP. As an example here let $q = 2$ and take w to be the probability function¹²

$$w = 4^{-1}(w_{\langle \frac{1}{2}, \frac{1}{2}, 0, 0 \rangle} + w_{\langle \frac{1}{2}, 0, \frac{1}{2}, 0 \rangle} + w_{\langle 0, \frac{1}{2}, 0, \frac{1}{2} \rangle} + w_{\langle 0, 0, \frac{1}{2}, \frac{1}{2} \rangle}).$$

Then w satisfies Ex and Px. However for $\theta = (P_1(a_1) \wedge \neg P_1(a_2))$, $\theta' = (P_2(a_3) \wedge \neg P_2(a_4))$, a straightforward calculation shows that

$$w(\theta | \theta') = 0 < w(\theta) = \frac{1}{8}.$$

Hence CP fails for this function.

A second argument for restricting attention here to probability functions satisfying ULi (equivalently to probability functions on L_∞) is that without it the lack of available predicates from which to form θ' from θ becomes a significant nuisance factor. Given this, and the fact that the main interest in Inductive Logic is in probability functions satisfying Ax, we shall henceforth limit our attention to probability functions satisfying ULi with Ax.

The Strict Counterpart Principle

In the previous section we considered the version of the Counterpart Principle, (3), asserting that the evidence θ' does not decrease the probability of θ . We now make some observations on when we can, and cannot, assert that θ' strictly increases the probability of θ , what one might call the ‘Strict Counterpart Principle’.

In fact we can never have this for *all* θ since for θ a tautology or contradiction θ' will have the same status and (3) will simply give equality. In this section we

¹²It is straightforward to see that convex combinations of probability functions also satisfy (P1-3) and hence are themselves probability functions.

shall look at some non-tautologous non-contradictory θ for which all probability functions satisfying ULi with Ax fail to give a strict inequality. First however we shall look at the other end of the spectrum, probability functions satisfying ULi with Ax which *fail* the strict version CP for *all* θ .

A particular class of well understood functions (see for example [11], [22], [24]) satisfying ULi with Ax, and so CP, are those which satisfy:

Weak Irrelevance Principle (WIP)

If $\theta, \phi \in SL$ have no constant or predicate symbols in common then

$$w(\theta | \phi) = w(\theta).$$

Clearly WIP implies that CP always holds with equality. For probability functions satisfying WIP we have a precise characterization which we now explain, in part because this notation will be required later.

Let \mathbb{B} be the set of infinite sequences

$$\bar{p} = \langle p_0, p_1, p_2, p_3, \dots \rangle$$

of reals such that $p_0 \geq 0, p_1 \geq p_2 \geq p_3 \geq \dots \geq 0$ and

$$\sum_{i=0}^{\infty} p_i = 1.$$

For $\bar{p} \in \mathbb{B}$ and $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, 2^q\}$ let $R_{\bar{p},n} = 1 - \sum_{j=1}^n p_j$ and designate

$$f(p) = \left\langle 2^{-q}R_{\bar{p},n} + \sum_{f(j)=1} p_j, 2^{-q}R_{\bar{p},n} + \sum_{f(j)=2} p_j, \dots, 2^{-q}R_{\bar{p},n} + \sum_{f(j)=2^q} p_j \right\rangle \in \mathbb{D}_{2^q}.$$

Now let $u_n^{\bar{p},L}$ be the probability function defined by

$$u_n^{\bar{p},L} = 2^{-nq} \sum_f w_{f(\bar{p})}$$

where the f range over all functions $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, 2^q\}$ and for $\theta \in QFSL$ define

$$u^{\bar{p},L}(\theta) = \lim_{n \rightarrow \infty} u_n^{\bar{p},L}(\theta).$$

This limit exists and by Theorem 1 $u^{\bar{p},L}$ extends to a probability function on L . The fact that the $u_n^{\bar{p},L}$ satisfy Ex and Ax carries over to $u^{\bar{p},L}$, indeed as we vary \mathcal{L} the $u^{\bar{p},L}$ form a language invariant family so $u^{\bar{p},L}$ satisfies ULi with Ax.

Notice that if $p_{n+1} = 0$ in \bar{p} then $u^{\bar{p},L} = u_n^{\bar{p},L}$. In particular for $0 \leq \delta \leq 1$ and $\bar{p} = \langle 1 - \delta, \delta, 0, 0, \dots \rangle$,

$$w_L^\delta = u^{\bar{p},L} = u_1^{\bar{p},L},$$

so these $u^{\bar{p},L}$ include the Nix-Paris Continuum.

A generalization (to polyadic languages) of the following theorem is proved in [22]

Theorem 4. *The $u^{\bar{p},L}$ are exactly the probability functions on L satisfying ULi with Ax and WIP.*

This theorem then provides a family of probability functions which have equality in CP (in the presence of ULi with Ax) for *all* $\theta \in SL$. We would conjecture that conversely these $u^{\bar{p},L}$ are the only probability functions with this property.

We now turn to look at non-tautologous non-contradictory sentences θ which guarantee equality in CP for *any* probability function satisfying ULi with Ax. To describe these we first need to introduce some more notation.

For $\theta \in QFSL$, let $f_\theta(\tilde{n})$ denote the number of state descriptions with spectrum $\tilde{n} = \{n_1, n_2, \dots, n_{2^q}\}$ appearing in the Disjunctive Normal Form of θ ¹³. Note that for any probability function w satisfying Ax and any sentence θ ,

$$w(\theta) = \sum_{\tilde{n}} f_\theta(\tilde{n})w(\tilde{n})$$

where $w(\tilde{n})$ is the value of w on some/any state description with spectrum \tilde{n} .

The following lemma appears in [23] but for completeness we include a proof here.

Lemma 5. *Let $\theta \in QFSL$ be such that for any probability function w on L satisfying Ax, $w(\theta) = c$, equivalently*

$$\sum_{\tilde{n}} f_\theta(\tilde{n})w(\tilde{n}) = c, \tag{4}$$

for some constant c . Then for each \tilde{n} , $f_\theta(\tilde{n}) = cf_{\neg}(\tilde{n})$.

Proof. Given reals $s_1, s_2, \dots, s_{2^q} \geq 0$, and not all zero, let $v_{\vec{s}}$ be the probability function on L such that

$$v_{\vec{s}}(\tilde{n}) = (2^q!)^{-1} \sum_{\sigma} s_{\sigma(1)}^{n_1} s_{\sigma(2)}^{n_2} \dots s_{\sigma(2^q)}^{n_{2^q}} (s_1 + s_2 + \dots + s_{2^q})^{-m}$$

¹³For some fixed set of constants which includes all the constants mentioned in θ , though the particular fixed set is not important in what follows.

where σ ranges over all permutations of $1, 2, \dots, 2^q$. Then $v_{\bar{s}}$ satisfies Ax and (4) together with the fact that $v_{\bar{s}}(\top) = 1$ gives that

$$\begin{aligned} \sum_{\tilde{n}} f_{\theta}(\tilde{n})(2^q!)^{-1} \sum_{\sigma} s_{\sigma(1)}^{n_1} s_{\sigma(2)}^{n_2} \dots s_{\sigma(2^q)}^{n_{2^q}} &= cv_{\bar{s}}(\top)(s_1 + s_2 + \dots + s_{2^q})^m \\ &= c \sum_{\tilde{n}} f_{\top}(\tilde{n})(2^q!)^{-1} \sum_{\sigma} s_{\sigma(1)}^{n_1} s_{\sigma(2)}^{n_2} \dots s_{\sigma(2^q)}^{n_{2^q}}. \end{aligned}$$

Since we can take each s_i to be algebraically independent this is only possible if the coefficients of $s_1^{n_1} s_2^{n_2} \dots s_{2^q}^{n_{2^q}}$ on both sides agree, from which the result follows. \square

We shall refer to a $\theta \in QFSL$ such that $w(\theta) = c$ for all probability functions w on L satisfying Ax as being of *constant type*. Notice that in this case c must be rational with denominator (when in lowest form) which divides all the $f_{\top}(\tilde{n})$.

As Lemma 5 is stated it seems possible that since the definition of a constant type sentence depends on the overlying language θ could be of constant type for L but not for L' even though $\theta \in SL \cap SL'$. The next lemma shows that this is not the case.

Lemma 6. *Suppose that $\theta \in SL \cap SL'$ and $w(\theta) = c$ for all probability functions w on L satisfying Ax. Then $w'(\theta) = c$ for all probability functions w' on L' satisfying Ax.*

Proof. It is enough to check this in the cases where $L \subset L'$ and where $L' \subset L$. In the former case the result is immediate since $w' \upharpoonright SL$ (w' restricted to SL) still satisfies Ax and $w' \upharpoonright SL(\theta) = w'(\theta)$.

So assume that $L' \subset L$. Then by Theorem 33.1 of [23] there is $\lambda \geq 0$ and probability functions w_1, w_2 on L satisfying Ax such that

$$w = (1 + \lambda)w_1 \upharpoonright SL' - \lambda w_2 \upharpoonright SL'. \quad (5)$$

Hence

$$\begin{aligned} w(\theta) &= (1 + \lambda)w_1 \upharpoonright SL'(\theta) - \lambda w_2 \upharpoonright SL'(\theta) \\ &= (1 + \lambda)w_1(\theta) - \lambda w_2(\theta) \\ &= (1 + \lambda)c - \lambda c, \quad \text{since } \theta \text{ is of constant type for } L, \\ &= c, \quad \text{as required.} \end{aligned}$$

\square

Theorem 7. For any $\theta \in QFSL$ of constant type and any $\phi \in QFSL$ that has no predicate or constant symbols in common with θ ,

$$w(\theta | \phi) = w(\theta)$$

for all w satisfying Ax.

Proof. Suppose that θ and ϕ have no constant or predicate symbols in common, and that

$$w(\theta) = m/n$$

for all probability functions w satisfying Ax. Clearly the result holds if $m = 0$ so assume that $m > 0$. Let L_1 to be the set (i.e. language) of all predicate symbols occurring in θ . By putting θ in Disjunctive Normal Form (DNF) we can express θ as a disjunction of state descriptions from L_1 . By Lemma 5

$$f_\theta(\tilde{n}) = (m/n)f_\top(\tilde{n})$$

for all spectra \tilde{n} of L_1 . For each spectrum \tilde{n} , let $r(\tilde{n}) = n^{-1}f_\top(\tilde{n})$ and partition the set of state descriptions of L_1 with spectrum \tilde{n} into n sets

$$A_1(\tilde{n}), A_2(\tilde{n}), \dots, A_n(\tilde{n}),$$

each containing $r(\tilde{n})$ state descriptions, so that the union of the first m of these sets constitute all the state descriptions in the DNF of θ with spectrum \tilde{n} .

Notice that for ξ, ψ state descriptions of L_1 (for the same constants) with spectrum \tilde{n} there is a permutation of atoms of L which sends $\xi \wedge \eta$ to $\psi \wedge \eta$ for any state description η of $L - L_1$ (with disjoint constants of course). Hence, by Ax, $w(\xi \wedge \eta) = w(\psi \wedge \eta)$, and in turn $w(\xi \wedge \phi) = w(\psi \wedge \phi)$ by taking the DNF equivalent of ϕ in $L - L_1$. So

$$\begin{aligned} w(\phi \wedge \theta) &= w\left(\phi \wedge \bigvee_{\tilde{n}} \bigvee_{j=1}^m \bigvee_{\psi \in A_j^{\tilde{n}}} \psi\right) \\ &= w\left(\bigvee_{\tilde{n}} \bigvee_{j=1}^m \bigvee_{\psi \in A_j^{\tilde{n}}} (\psi \wedge \phi)\right) \\ &= \sum_{\tilde{n}} \sum_{j=1}^m \sum_{\psi \in A_j^{\tilde{n}}} w(\psi \wedge \phi) \\ &= \sum_{\tilde{n}} m r(\tilde{n}) w(\psi_{\tilde{n}} \wedge \phi) \end{aligned}$$

for some/any state description $\psi_{\tilde{n}}$ of L_1 (for the same constants) with spectrum \tilde{n} . Exactly similarly by replacing θ by \top ,

$$w(\phi) = \sum_{\tilde{n}} nr(\tilde{n})w(\psi_{\tilde{n}} \wedge \phi) = (n/m)w(\phi \wedge \theta)$$

so since $w(\theta) = m/n$ the required identity follows. \square

It is worth remarking here that this theorem can, with some more effort, be proved even for $\phi \in SL$ (see [23]) though we will not need that stronger version here.

From Theorem 7 we have the following corollary.

Corollary 8. *For any $\theta \in QFSL$ of constant type and any θ' obtained by replacing all constant and predicate symbols in θ by new ones,*

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

for all w satisfying Ax.

The converse to Corollary 8 is easily shown.

Proposition 9. *Suppose that $\theta, \theta' \in QFSL$ with θ' the result of replacing all constant and predicate symbols in θ by new ones and*

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

for all w satisfying Ax. Then θ is of the constant type.

Proof. Let w_1, w_2 be distinct probability functions satisfying Ax. Then the probability function $2^{-1}(w_1 + w_2)$ also satisfies Ax and so by assumption,

$$\begin{aligned} 2^{-1}(w_1 + w_2)(\theta \wedge \theta') &= 2^{-1}(w_1 + w_2)(\theta)2^{-1}(w_1 + w_2)(\theta') \\ &= (2^{-1}(w_1 + w_2)(\theta))^2 \end{aligned}$$

since by Ax $w(\theta) = w(\theta')$. Multiplying out and re-arranging we get

$$2w_1(\theta \wedge \theta') + 2w_2(\theta \wedge \theta') = w_1(\theta)^2 + 2w_1(\theta)w_2(\theta) + w_2(\theta)^2.$$

Using the assumption this gives that

$$2w_1(\theta)^2 + 2w_2(\theta)^2 = w_1(\theta)^2 + 2w_1(\theta)w_2(\theta) + w_2(\theta)^2$$

and by re-arranging

$$(w_1(\theta) - w_2(\theta))^2 = 0.$$

Hence $w_1(\theta) = w_2(\theta)$ as required. \square

Having seen a class of sentences for which equality always holds in the statement of CP, we turn to consider a case in which strict inequality holds for all non-constant $\theta \in QFSL$. In order to do so we recall the following special case of a theorem (Theorem 1) from [14].

Theorem 10. *Any probability function w on L satisfying ULi with Ax can be represented as an integral*

$$w = \int_{\mathbb{B}} u^{\bar{p},L} d\mu \quad (6)$$

for some measure μ on the Borel subsets of \mathbb{B} .

Conversely any such function defined in this way satisfies ULi with Ax.¹⁴

Theorem 11. *For a probability function $w = \int_{\mathbb{B}} u^{\bar{p},L} d\mu$, if every point in \mathbb{B} is a support¹⁵ point of μ then strict inequality holds in CP whenever $\theta \in QFSL$ is not of the constant type.*

Proof. Assume that w can be expressed in this way and let $\theta, \theta' \in QFSL$ be as in the statement of CP. Then since the $u^{\bar{p},L}$ satisfy WIP,

$$\begin{aligned} w(\theta \wedge \theta') - w(\theta)^2 &= \int_{\mathbb{B}} u^{\bar{p},L}(\theta \wedge \theta') d\mu(\bar{p}) - \left(\int_{\mathbb{B}} u^{\bar{q},L}(\theta) d\mu(\bar{q}) \right)^2 \\ &= \int_{\mathbb{B}} u^{\bar{p},L}(\theta)^2 d\mu(\bar{p}) - \left(\int_{\mathbb{B}} u^{\bar{q},L}(\theta) d\mu(\bar{q}) \right)^2 \\ &= \int_{\mathbb{B}} \left(u^{\bar{p},L}(\theta) - \int_{\mathbb{B}} u^{\bar{q},L}(\theta) d\mu(\bar{q}) \right)^2 d\mu(\bar{p}) \geq 0. \end{aligned}$$

Since the support of μ is all of \mathbb{B} the only way we can have $w(\theta | \theta') = w(\theta)$ is if

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

for all $\bar{p} \in \mathbb{B}$. In other words $u^{\bar{p},L}(\theta)$ must be constant for all $\bar{p} \in \mathbb{B}$.

By a result in [23, Chapter 33] any probability function w on L satisfying Ax is of the form

$$w = (\lambda + 1) \int_{\mathbb{B}} u^{\bar{p},L} d\mu_1(\bar{p}) - \lambda \int_{\mathbb{B}} u^{\bar{p},L} d\mu_2(\bar{p})$$

for some $0 \leq \lambda$ and measures μ_1, μ_2 on \mathbb{B} . Hence if all the $u^{\bar{p},L}(\theta)$ are constant then so too are all $w(\theta)$ for w satisfying Ax. In other words θ is of the constant type. \square

¹⁴Notice that the ‘building block functions’ here, i.e. the $u^{\bar{p},L}$, are precisely the probability functions satisfying ULi with Ax and WIP.

¹⁵Recall that a point $\vec{e} \in \mathbb{B}$ is in the *support* of μ if $\mu(B) > 0$ for all open subsets B of \mathbb{B} containing \vec{e} .

The conditions given in this theorem which ensure that w satisfies CP with strict inequality for all non constant type quantifier free sentences can be shown to hold for Carnap's Continuum of Inductive Methods c_λ when $0 < \lambda < \infty$, thus ensuring that these c_λ satisfy this strong version of CP (for quantifier free sentences). However showing this appears to be quite involved and in general we currently have little insight into when these conditions hold for particular probability functions (unlike the situation with the de Finetti's Representation).

It might have been hoped at this point that any probability function w satisfying ULi with Ax would either satisfy WIP, and so *never* give strict inequality in CP, or else not satisfy WIP and *always* give strict inequality in CP whenever $\theta \in QFSL$ was not of the constant type. Unfortunately as the following example shows the situation is not as simple as that.

Let L be the language with just two predicate symbols, i.e. $q = 2$. Then for a state description θ with spectrum $\{3, 1, 0, 0\}$ or $\{2, 2, 0, 0\}$ the mapping $\delta \mapsto w^\delta(\theta)$ has a maximum point in $(0, 1)$. So there are $0 < \nu < \tau < 1$ such that $w^\nu(\theta) = w^\tau(\theta)$ and in consequence the probability function $w = (w^\nu + w^\tau)/2$ has the property that $w(\theta | \theta') = w(\theta)$. However θ is not of the constant type and w does not satisfy WIP.

Conclusions

This paper argues that the Counterpart Principle of Analogy has intuitive appeal and gives a proof that it holds for all probability functions satisfying Unary Language Invariance, another appealing principle in our opinion. In particular then the Counterpart Principle is satisfied by both Carnap's Continuum of Inductive Methods and the Nix-Paris Continuum. This also means that it is consistent with Atom Exchangeability, which makes it a rather different analogy principle to those considered previously.

The paper also investigates when the inequality in the Counterpart Principle is strict. In this case it is shown that there are probability functions satisfying Unary Language Invariance with Atom Exchangeability that never give strict inequality for any sentence (amongst them the w^δ of the Nix-Paris Continuum) and there are non-tautologous non-contradictory 'constant' sentences which never give a strict inequality for any probability function satisfying Atom Exchangeability. Whilst the general situation still begs clarification it can be shown that for $0 < \lambda < \infty$ the c_λ of Carnap's Continuum of Inductive Methods do give the strict inequality on quantifier free sentences for all except the 'constant' sentences.

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., The Aim of Inductive Logic, in *Logic, Methodology and Philosophy of Science*, Eds. E.Nagel, P.Suppes & A.Tarski, Stanford University Press, Stanford, California, 1962, pp303-318.
- [4] 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, Berkeley and Los Angeles, 1980, pp7-155.
- [5] Carnap, R. & Stegmüller, W., *Induktive Logik und Wahrscheinlichkeit*, Springer Verlag, Wien, 1959.
- [6] Festa, R., Analogy and Exchangeability in Predictive Inferences, *Erkenntnis*, 1996, **45**:229-252.
- [7] de Finetti, B., *Theory of Probability, Volume 1*, Wiley, New York, 1974.
- [8] Gaifman, H., Concerning measures on first order calculi, *Israel Journal of Mathematics*, 1964, **2**:1-18.
- [9] Gaifman, H., Applications of de Finetti's Theorem to Inductive Logic, in *Studies in Inductive Logic and Probability, Volume I*, eds. R.Carnap & R.C.Jeffrey, University of California Press, Berkeley and Los Angeles, 1971, pp235-251.
- [10] Hill, A.J. & Paris, J.B., Reasoning by Analogy in Inductive Logic, *The Logica Yearbook 2011*, eds. M.Peliš & V.Punčochář, College Publications, London, 2012, pp63-76.
- [11] Hill, M.J., Paris, J.B. & Wilmers, G.M., Some observations on induction in predicate probabilistic reasoning, *Journal of Philosophical Logic*, 2002, **31**:43-75.
- [12] 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, pp711-738.
- [13] Kuipers, T.A.F., Two types of Inductive Analogy by Similarity, *Erkenntnis*, 1984, **21**:63-87.

- [14] Landes, J., Vencovská, A. & Paris, J.B., A Characterization of the Language Invariant Families satisfying Spectrum Exchangeability in Polyadic Inductive Logic, *Annals of Pure and Applied Logic*, 2010, **161**:800-811.
- [15] Landes, J., Paris, J.B. & Vencovská, A survey of some recent results on Spectrum Exchangeability in Polyadic Inductive Logic, *Synthese*, 2011, **181**(Supplement 1), pp19-47.
- [16] Maher, P., Probabilities for multiple properties: The models of Hesse, Carnap and Kemeny, *Erkenntnis*, 2001, **55**:183-216.
- [17] Maher, P., A Conception of Inductive Logic, *Philosophy of Science*, 2006, **73**:513-520.
- [18] di Maio, M.C., Predictive Probability and Analogy by Similarity, *Erkenntnis*, 1995, **43**(3):369-394.
- [19] Niiniluoto, I., Analogy and Inductive Logic, *Erkenntnis*, 1981, **16**:1-34.
- [20] Nix, C.J. & Paris, J.B., A Continuum of Inductive Methods arising from a Generalized Principle of Instantial Relevance, *Journal of Philosophical Logic*, 2006, **35**(1):83-115.
- [21] Paris, J.B., Pure Inductive Logic, in *The Continuum Companion to Philosophical Logic*, Eds. L.Horsten & R.Pettigrew, Continuum International Publishing Group, London, 2011, pp428-449.
- [22] Paris, J.B. & Vencovská, A., A Note on Irrelevance in Inductive Logic, *Journal of Philosophical Logic*, 2010, **40**(3):357-370.
- [23] Paris, J.B. & Vencovská, A., *Pure Inductive Logic*. To appear in the Association of Symbolic Logic series *Perspectives in Logic*, Ed. M.Rathjen, Cambridge University Press, 2013.
- [24] Paris, J.B. & Waterhouse, P., Atom Exchangeability and Instantial Relevance, *Journal of Philosophical Logic*, 2009, **38**(3):313-332.
- [25] Polya, G., *Induction and Analogy in Mathematics, Volume I of Mathematics and Plausible Reasoning*, Geoffrey Cumberlege, Oxford University Press, 1954.
- [26] Polya, G., *Patterns of Plausible Inference, Volume II of Mathematics and Plausible Reasoning*, Geoffrey Cumberlege, Oxford University Press, 1954.

- [27] Romeijn, J-W., Analogical Predictions for Explicit Similarity, *Erkenntnis*, 2006, **64**(2):253-280.
- [28] Skyrms, B., Analogy by Similarity in Hyper-Carnapian Inductive Logic, in *Philosophical Problems of the Internal and External Worlds*, J.Earman, A.I.Janis, G.Massey & N.Rescher, eds., University of Pittsburgh Press, 1993, pp273-282.