Combining Analogical Support in Pure Inductive Logic

Paris, J.B. and Vencovska, A.

2017

MIMS EPrint: **2017.3**

Manchester Institute for Mathematical Sciences

School of Mathematics

The University of Manchester

Combining Analogical Support in Pure Inductive Logic

J.B.Paris[*]and A.Vencovská[†]
School of Mathematics
The University of Manchester
Manchester M13 9PL

jeff.paris@manchester.ac.uk,   alena.vencovska@manchester.ac.uk

June 9, 2016

Abstract

We investigate the relative probabilistic support afforded by the combination of two analogies based on possibly different, structural similarity (as opposed to e.g. shared predicates) within the context of Pure Inductive Logic and under the assumption of Language Invariance. We show that whilst repeated analogies grounded on the same structural similarity only strengthen the probabilistic support this need not be the case when combining analogies based on different structural similarities. That is, two analogies may provide less support than each would individually.

Keywords: The Counterpart Principle, Structural Similarity, Analogy, Inductive Logic, Logical Probability, Rationality, Uncertain Reasoning.

1

Introduction

Suppose that I am considering how likely it is that my son would enjoy a visit to the cinema to see *The Sound of Music*. Thinking about it I recall that last year he did enjoy seeing *Toy Story* which somewhat enhances my belief that he will also enjoy *The Sound of Music*. I then remember my aunt telling me how much she enjoyed it. Should this now further increase the probability I would give to my son liking it?

Here is an example of two facts which individually seem to provide analogical support for my son liking *The Sound of Music* yet in combination they seem to be pulling in different directions and possibly canceling each other out. (Like enjoying curry and rhubarb crumble but not both at the same time!)

The plan in this paper is to investigate this issue of combining analogical support (at least up to two such supports) within the context of Pure Inductive Logic (PIL for short). Of course since this version of Carnapian Inductive Logic considers the assignment of rational, or logical, probability in the absence of any intended interpretation of the language on closer scrutiny the above example hardly seems relevant (since we already know a great deal about films, aunts, etc. so they are very far indeed from being uninterpreted). Nevertheless it still seems to us interesting to consider this question 'in vacuo', in other words as a nascent artificial agent might do.

Within philosophy, and similarly in AI and psychology, there is a considerable literature on analogy, see Bartha's [1], [2] for an excellent overview. Bartha however explicitly avoids considering approaches to analogical reasoning within the general framework of Carnap's Inductive Logic, as found for example in [3], [6], [7], [8], [11], [23], [24], [25], [26], [27], [28], [29], [30], [35], [36], [38], [39], [40].[1] It is within this framework, more specifically within PIL and directly following on from the earlier [18], [19], [20], [21], [33], that this present paper is set. Working within such a mathematical setting has the advantage of precision and permanence, a correct theorem cannot be denied only at worst put aside as irrelevant, though we will argue later that results in this formal backwater do have relevance within the wider ambit.

Notation and Context

The context of this paper is PIL as explained, for example, in [31] and [33]. Thus we have a language L with relation symbols $R_1, R_2, \ldots, R_q$, say of finite arities

---

[1]See [17], [19] for some discussion of this literature.

$r_1, r_2, \ldots, r_q$ respectively, and constants a_n for $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$, and no function symbols nor equality.[2] Let SL denote the set of first order sentences of this language L.

We are interested in picking a rational *probability function* w on SL, i.e. a function $w : SL \to [0, 1]$ such that for $\theta, \phi, \exists x\, \psi(x) \in SL$

(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) $w(\exists x\, \psi(x)) = \lim_{m \to \infty} w(\bigvee_{i=1}^{m} \psi(a_i))$,

where *rational* is customarily identified with w satisfying certain arguably rational principles. Whilst there is currently no clear consensus on what these principles should be one that is very widely adopted is

The Principle of Constant Exchangeability, Ex.

A probability function w on SL satisfies Constant Exchangeability if, for any permutation σ of $1, 2, \ldots$ and $\theta(a_1, \ldots, a_m) \in SL$,

$$w(\theta(a_{\sigma(1)}, \ldots, a_{\sigma(m)})) = w(\theta(a_1, \ldots, a_m)).$$

All the probability functions considered in this paper will be assumed to satisfy Ex.

A second similar principle is

The Principle of Predicate Exchangeability, Px

If R_i, R_j are relation symbols of L_q, of the same arity, then for $\theta \in SL$,

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

where θ' is the result of transposing R_i, R_j throughout θ.

Px necessarily differs from Ex as far as our default language L is concerned in that L has infinitely many constant symbols but only finitely many relation symbols.[3] If w satisfying Ex+Px can be extended to a probability function w_∞ on the sentences

---

[2]We identify L with the set $\{R_1, R_2, \ldots, R_q\}$.

[3]In the study of PIL there are good reasons for this differing treatment, in particular almost all principles currently being considered are adequately captured in languages with only finitely many relation symbols, unlike the case for constants.

of a language L_∞ with infinitely many relation symbols of each arity and continue to satisfy Ex+Px we say that w satisfies *Language Invariance*, Li. This is a stronger condition on w than simply satisfying Ex+Px nevertheless it holds widely for the main probability functions considered in this subject, for example Carnap's Continuum of Inductive Methods. It is also arguably rational on the grounds that the probability assigned to a sentence θ should surely not depend on the presence or otherwise of relation symbols in the overlying language which are not mentioned in θ. This amounts to the requirement that a rational probability function on a language L should be extendible to a rational probability function on any larger language, and once Px is taken as a condition for rationality this is equivalent to Li by a sequential compactness argument (see [32] for more details).

The Results

In [20] a further putative rationality principle was proposed based on 'probabilistic analogical support by structural similarity': [4]

The Counterpart Principle, CP

Let $\theta, \theta' \in SL$ be such that θ' is the result of replacing some constant/relation symbols in θ by new constant/relation symbols not occurring in θ. Then[5]

$$w(\theta \,|\, \theta') \geq w(\theta).$$

For example in the case where R_1 was unary and R_2, R_3 binary and

$$\theta = \neg R_1(a_1) \wedge \exists x \, R_2(x, a_2)$$

we could have that

$$\theta' = \neg R_1(a_3) \wedge \exists x \, R_3(x, a_2)$$

and this instance of CP would give

$$w(\neg R_1(a_1) \wedge \exists x \, R_2(x, a_2) \,|\, \neg R_1(a_3) \wedge \exists x \, R_3(x, a_2)) \geq w(\neg R_1(a_1) \wedge \exists x \, R_2(x, a_2)).$$

We are thinking of θ and θ' in the statement of CP as capturing what is meant by *structural similarity*, that they have exactly the same syntactic form, only some

---

[4]This underlying conception of analogy is similar to that employed in *Structure-Mapping Theory* in AI, see [14].

[5]Throughout we avoid any problems of conditioning on sentences with zero probability by identifying, for example, $w(\phi \,|\, \psi) \geq w(\eta \,|\, \zeta)$ with $w(\phi \wedge \psi) \cdot w(\zeta) \geq w(\eta \wedge \zeta) \cdot w(\psi)$.

names have changed. For example, purely in terms of the syntax and in the absence of any interpretation my aunt enjoying *The Sound of Music* at Christmas is structurally similar to my son enjoying *The Sound of Music* on his birthday. For such structurally similar θ, θ' just Ex+Px force that they must get the same probability. The Counterpart Principle adds to this formal connection by asserting that there is also a material connection in that conditioning on θ' enhances (or at least does not diminish) the probability of θ.

Relating CP to the common format of analogical reasoning within philosophy as a whole Bartha, [1], gives a Candidate Analogical Inference Rule (R) saying, in short, that having a mapping $*$ from a source domain $\mathcal{S}$ to a target domain $\mathcal{T}$ the plausibility of a proposition Q^* holding in $\mathcal{T}$ given that Q holds in $\mathcal{S}$ should be stronger the more features θ, θ^* the domains have in common as opposed to features on which they differ. Within this template CP corresponds to the null case where any evidence of matching/dismatching features is absent (a case that Bartha does not treat). In fact as we shall explain in the penultimate section allowing in even one such matching feature can, rather unexpectedly, destroy the analogical support in the way we are measuring it.

The general pattern of support by analogy that CP endeavours to capture seems to be rather common in our everyday lives, and in science. For example learning that every natural number can be proved to be the sum of four squares might well cause one to guess that every natural number can also be proved to be the sum of nine cubes. From such familiarity one might then feel that there was possibly a case to argue for the rationality of CP. In fact we do not really need to do so since by the following theorem from [20] it is actually inherited from Li.

Theorem 1. *Li implies CP.*

In a sense one might say that this theorem provides one answer to the question raised by M.Hesse in 1966, [16], as to what is the philosophical justification for analogical reasoning. For whenever we reason it is necessary to simply discard almost everything we know - there is just too much of it to fully incorporate. We must ring fence a tiny fraction of it that we consider relevant and reason simply on the basis of that knowledge. Theorem 1 tells us that if all we consider relevant is our knowledge that θ' holds and we assign probabilities rationally, in the sense of satisfying Li (and Ex) then the probability of θ will be enhanced.[6] From this viewpoint the derived analogical support arises through adopting Li.

---

[6]Or at least not diminished – the conditions for merely equality in CP are rather messy but in our view sufficiently 'unnatural' as to offer no serious grounds on which to criticize CP, see [20] for more details.

Still we should emphasize here that what is 'derived' is an increased belief in θ within the ring fence. Whether or not this has any relevance to beliefs in the wider world outside the ring fence depends on how far we can accept the ring fence assumption.[7] It should also be made clear that we are not claiming that this increased belief equates with increased probability of being 'true' in any objective sense. But what it might be argued to provide is plausibility, and in turn the impetus to attempt to confirm a truth, though no more than that. For example the discovery that every natural number is the sum of $4 = 2^2$ squares might equally have encouraged mathematicians to try to prove that every natural number is the sum of $2^3 = 8$ cubes rather than $3^2 = 9$ cubes.

As a further development of the idea of structural similarity as embodied in CP the following result is shown in [33]:

Theorem 2. *Suppose that $\theta, \theta', \theta'' \in SL$ are such that θ' is the result of replacing some constant/relation symbols in θ by new constant/relation symbols not occurring in θ and similarly θ'' is the result of replacing some constant/relation symbols in θ' by new constant/relation symbols not occurring in θ or θ'. Then for w satisfying Li,[8]*

$$w(\theta \,|\, \theta') \geq w(\theta \,|\, \theta''). \tag{1}$$

Carrying on from the above example then this gives that

$$w(\neg R_1(a_1) \wedge \exists x \, R_2(x, a_2) \,|\, \neg R_1(a_3) \wedge \exists x \, R_3(x, a_2))$$
$$\geq \; w(\neg R_1(a_1) \wedge \exists x \, R_2(x, a_2) \,|\, \neg R_1(a_3) \wedge \exists x \, R_3(x, a_4)).$$

For $\theta, \theta', \theta''$ as in this theorem there is a clear sense in which θ' is structurally at least as similar to θ as θ'' is. In other words there is a sense of distance, or degree of similarity, the nearer θ' is to θ the more analogical support it provides. In summary then we could say that Theorem 1 tells us that θ' provides 'analogical support by structural similarity' for θ under the assumption that w satisfies Li and Theorem 2 tells us that the closer the analogy (i.e. structural similarity) the stronger the analogical support.

Theorem 2 raises the question of what the effect is of combining multiple such analogies and more generally what is the algebra of analogical support by similarity

---

[7]Possibly this explains the force of analogy in mathematics and science where the outside world according to our current view is so structured and lawlike.

[8]With nothing more than complicating the notation we could also add to the conditioning sentences a sentence ψ provided that none of the relation and constant symbols which replace or are replaced in the passage from θ to θ' to θ'' appear in ψ. This remark also applies to Theorem 3.

in the presence of Li? We will certainly not answer that question in this paper but will settle for elucidating the situation when 'multiple' is weakened to 'two, at most'.

In order to state our results, which will be proved in full in the next section, it will be useful to sketch an artifice in the proof of Theorem 2 as given in [33] because it enables us to reduce these questions to a particularly simple form. Leaving implicit any constant and relation symbols common to them all, we can write these θ, θ' and θ'' as

$$\begin{aligned} \theta &= \theta(\vec{B}_1, \vec{C}_1, \vec{D}_1), \\ \theta' &= \theta(\vec{B}_1, \vec{C}_2, \vec{D}_2), \\ \theta'' &= \theta(\vec{B}_2, \vec{C}_2, \vec{D}_3), \end{aligned}$$

where the $\vec{B}_1, \vec{B}_2, \vec{C}_1, \vec{C}_2, \vec{D}_1, \vec{D}_2, \vec{D}_3$ are entirely disjoint blocks of distinct constant and/or relation symbols and the $\vec{B}_1, \vec{B}_2$ are matching in terms of the positions of constant and relation symbols, and similarly for the $\vec{C}, \vec{D}$. So for example if $\vec{B}_1 = \langle a_1, R_1, R_2 \rangle$ with R_1 unary and R_2 ternary then $\vec{B}_2$ must similarly be of the form $\langle a_k, R_m, R_g \rangle$ with R_m unary, R_g ternary.

Let the probability function w on SL satisfy Li and as explained above let w_∞ be the extension of w to SL_∞ satisfying Ex+Px. Let $\vec{B}_n$ for $n > 2$ be blocks of totally new relation and constant symbols from L_∞ matching $\vec{B}_1$ and similarly produce $\vec{C}_n, \vec{D}_n$ matching $\vec{C}_1, \vec{D}_1$ respectively, so that no constant or relation symbol appears in more than one block of any sort. Notice that since w_∞ extends w and satisfies Ex+Px to show that $w(\theta \mid \theta') \geq w(\theta \mid \theta'')$ it is enough to show that

$$w_\infty(\theta(\vec{B}_1, \vec{C}_2, \vec{D}_2) \mid \theta(\vec{B}_1, \vec{C}_3, \vec{D}_3)) \geq w_\infty(\theta(\vec{B}_1, \vec{C}_2, \vec{D}_2) \mid \theta(\vec{B}_3, \vec{C}_4, \vec{D}_4)). \quad (2)$$

Let $\mathcal{L}$ be the language with a single binary relation symbol R. Define a probability function v on $S\mathcal{L}$ by

$$v(R(a_i, a_j)) = w_\infty(\theta(\vec{B}_i, \vec{C}_j, \vec{D}_j))$$

and more generally

$$v\left(\bigwedge_{i,j \leq n} R(a_i, a_j)^{\epsilon_{i,j}}\right) = w_\infty\left(\bigwedge_{i,j \leq n} \theta(\vec{B}_i, \vec{C}_j, \vec{D}_j)^{\epsilon_{i,j}}\right)$$

where the $\epsilon_{i,j} \in \{0, 1\}$ and $R^1 = R, R^0 = \neg R$, etc. [By a theorem of Gaifman, see [12] (or [33]), v extends uniquely to a probability function on $S\mathcal{L}$ satisfying Ex.]

Referring to [33, page 187] for the details it can now be shown that for v satisfying Ex,

$$v(R(a_1, a_2) \mid R(a_1, a_3)) \geq v(R(a_1, a_2) \mid R(a_3, a_4)). \qquad (3)$$

But this is just another way of formulating the inequality (2).

From this sketch we see that to show that (1) holds for w satisfying Li it is enough to show that (3) holds for a probability function v on $S\mathcal{L}$ satisfying Ex.

We shall use the same method to investigate other simple cases of 'analogical support by (multiple) instances of structural similarity'. Note that any probability function on $S\mathcal{L}$ satisfying Ex also satisfies Li: for $\mathcal{L}$ has just one (binary) relation symbol and if v is a probability function on a language with at most one relation symbol of each arity which satisfies Ex then v trivially satisfies Px and also satisfies Li since we can just introduce relation symbols which for each arity are identical (to within the name). Hence showing above that (1) held for w satisfying Li was in fact *equivalent* to showing that (3) held for v on $S\mathcal{L}$ satisfying Ex; any v on $S\mathcal{L}$ satisfying Ex and failing (3) would have provided a counterexample to (1) satisfying Li.

This observation leads us to concentrate on deciding which inequalities must hold between

(a) $w(R(a_1, a_2))$

(b) $w(R(a_1, a_2) \mid R(a_3, a_4))$ (c) $w(R(a_1, a_2) \mid R(a_1, a_4))$

(d) $w(R(a_1, a_2) \mid R(a_3, a_4) \wedge R(a_5, a_6))$ (e) $w(R(a_1, a_2) \mid R(a_1, a_4) \wedge R(a_5, a_6))$

(f) $w(R(a_1, a_2) \mid R(a_1, a_4) \wedge R(a_1, a_6))$ (g) $w(R(a_1, a_2) \mid R(a_1, a_4) \wedge R(a_5, a_2))$

for w a probability function on $S\mathcal{L}$ satisfying Ex. In other words we are interested in comparing the degrees of support afforded by up two individually structurally similar sentences.[9] To illustrate these within a less formal context suppose that we read 'enjoys' for R, 'my son' for a_1, *The Sound of Music* for a_2 and so on. Then (a) would correspond, ceteris paribus, to the probability I would give to my son enjoying *The Sound of Music*, (c) to his enjoying it on my recalling that he enjoyed *Toy Story* and (g) to when I also recalled that my aunt enjoying the sound of music. Likewise (b) would correspond to the probability I would give to my son enjoying *The Sound of Music* just given that my work colleague enjoyed *Toy*

---

[9]This is not an exhaustive list of such conditionals, a point we shall return to later.

Story.[10] [Of course this informal example is very special because in the general case, as indicated in the earlier, R will correspond to a template for a sentence and the a_i blocks of constants and/or relations.]

The answers we obtain are given by the following figure 1 where a connection from x to y by upward sloping lines than means that y is always at least as large as x (and for some w strictly larger) whilst in unconnected cases the inequality can go either way. In each case, as above, establishing the inequality here provides a corresponding inequality for the more general case of multiple analogical support whilst a counterexample to the inequality holding itself provides a counterexample to the corresponding theorem for multiple analogical support.

So, with the above ring fence assumptions in place, (a) $\leq$ (b) prescribes that the probability I would give that my son will enjoy The Sound of Music should not be diminished when I recall his previously enjoying Toy Story. Similarly (d) $\leq$ (g) would prescribe one giving at least as much probability to the generalized Pell's equation $x^2 - 3y^2 = 6$ having a solution when judging (just!) on the basis of $x^2 - 3y^2 = 1$ and $x^2 - 10y^2 = 6$ having solutions than when judging instead on the basis of $x^2 - 5y^2 = 4$ and $x^2 - 2y^2 = 1$ having solutions.

fig.(i)
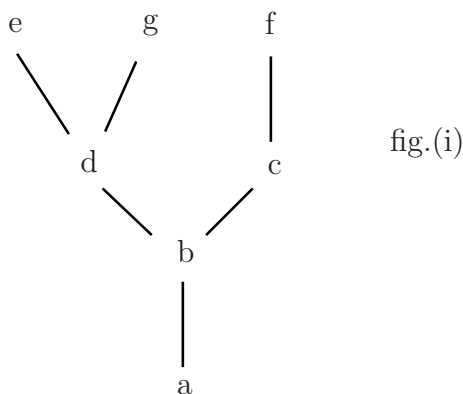
Each of these inequalities then corresponds to a 'Principle of Analogical Support by Structural Similarity' satisfied by a probability function satisfying Li. For example

---

[10] It is of course essential in these cases that one treats them purely as given without introducing any further knowledge or interpretation. Difficult as this is for anything but a nascent artificial agent, still it is the case that in everyday reasoning we do ring fence the available information between relevant and irrelevant. All we ask of our reader here is that they use ring fencing that includes only what is explicitly given.

(b) $\leq$ (e) gives that for w satisfying Li,

$$w(\exists x\, R_2(x, a_3) \,|\, \exists x\, R_3(x, a_5)) \leq w(\exists x\, R_2(x, a_3) \,|\, \exists x\, R_2(x, a_5) \wedge \exists x\, R_3(x, a_2))$$

whilst we cannot in general conclude that for such a w the instance

$$w(\exists x\, R_2(x, a_3) \,|\, \exists x\, R_3(x, a_5) \wedge \exists x\, R_4(x, a_2)) \leq w(\exists x\, R_2(x, a_3) \,|\, \exists x\, R_2(x, a_5))$$

of (d) $\leq$ (c) holds.

As we shall see for the most part the inequalities and incompatibilities in the above diagram will be straightforward to show. The main novelty in this paper is in proving Theorem 4, that (d) $\leq$ (g).

The Proofs

We now set about proving that the inequalities indicated in the above diagram are all that necessarily hold for w a probability function on $S\mathcal{L}$ satisfying Ex. We shall start by giving the positive results. In order to do this, as well as to later furnish some counterexamples, we need to recall (and slightly rejig) the Representation Theorem for polyadic probability functions as given in [33, Chapter 25] in the special case of the language $\mathcal{L}$ as above.

Let $D = (d_{i,j})$ be an $N \times N$ $\{0, 1\}$-matrix. Define a probability function w^D on $S\mathcal{L}$ by setting

$$w^D \left(\bigwedge_{i,j \leq n} R(a_i, a_j)^{\epsilon_{i,j}} \right)$$

to be the probability of (uniformly) randomly picking, with replacement, $h(1), h(2), \ldots, h(n)$ from $\{1, 2, \ldots, N\}$ such that for each $i, j \leq n$, $d_{h(i),h(j)} = \epsilon_{i,j}$. This uniquely determines a probability function on $S\mathcal{L}$ satisfying Ex. (For details see e.g. [33, Chapter 7]).

Clearly convex mixtures of these w^D also satisfy Ex. The converse (that any probability function satisfying Ex is a convex mixture of these w^D) does not hold *but* as shown in [33, Chapter 25] if we move to a suitable non-standard universe and take N to be a non-standard integer then any probability function w on $S\mathcal{L}$ satisfying Ex *is* a convex mixture (as an integral) of the standard parts $^{\circ}w^D$ of w^D from this non-standard universe (see [33, Theorem 25.1]). The important point to take away from this as far as this paper is concerned is that it is sometimes enough to check whether an inequality holds for these $^{\circ}w^D$ and since the mathematics is exactly the same it is just as good to check it for the w^D for standard N.

Given this observation, it turns out that that (d) $\leq$ (g) is a consequence of the following theorem whose rather messy proof is given in [34].

Theorem 3. *Let $(d_{i,j})$ be an $N \times N$ $\{0,1\}$-matrix such that $\sum_{i,j} d_{i,j} = T > 0$. Then*

$$\sum_{i,j} d_{i,j} A_i B_j \geq T^3 N^{-2}$$

where $A_i = \sum_r d_{i,r}$, $B_j = \sum_s d_{s,j}$.

Given this inequality we can now prove the main theorem of this paper which shows that (d) $\leq$ (g).

Theorem 4. *Let w be a probability function on the sentences of the binary language $\mathcal{L} = \{R\}$ satisfying Ex. Then*

$$w(R(a_1, a_2) \,|\, R(a_1, a_3) \wedge R(a_4, a_2)) \geq w(R(a_1, a_2) \,|\, R(a_3, a_4) \wedge R(a_5, a_6)).$$

Proof. For D an $N \times N$ $\{0,1\}$-matrix and $n \neq m$, $w^D(R(a_n, a_m))$ is the probability of picking $h(n), h(m) \in \{1, 2, \ldots, N\}$ such that $d_{h(n),h(m)} = 1$. In other words

$$w^D(R(a_n, a_m)) = N^{-2} \sum_{i,j} d_{i,j}.$$

Hence, since the choices $h(1), h(2), h(3), h(4), h(5), h(6)$ are all independent,

$$w^D(R(a_1, a_2) \wedge R(a_3, a_4) \wedge R(a_5, a_6)) = w^D(R(a_1, a_2))^3 = N^{-6} \left(\sum_{i,j} d_{i,j} \right)^3. \quad (4)$$

Likewise $w^D(R(a_1, a_2) \wedge R(a_1, a_3) \wedge R(a_4, a_2))$ is the probability of picking $h(1), h(2)$ such that $d_{h(1),h(2)} = 1$ and then (independently) picking $h(3), h(4)$ such that $d_{h(1),h(3)} = 1$ and $d_{h(4),h(2)} = 1$. For a particular choice of $h(1), h(2)$ these latter two probabilities are respective

$$N^{-1} \sum_s d_{h(1),s}, \quad N^{-1} \sum_r d_{r,h(2)}.$$

Thus the combined probability of these with $d_{h(1),h(2)} = 1$ is

$$N^{-2} \sum_{i,j} \left(d_{i,j} \left(N^{-1} \sum_s d_{i,s} \right) \left(N^{-1} \sum_r d_{r,j} \right) \right).$$

11

By Theorem 3 this is at least

$$N^{-6}\left(\sum_{i,j} d_{i,j}\right)^3$$

and combined with (4) we obtain that

$$w^D(R(a_1,a_2) \wedge R(a_1,a_3) \wedge R(a_4,a_2)) \geq w^D(R(a_1,a_2) \wedge R(a_3,a_4) \wedge R(a_5,a_6)).$$

Being a linear inequality this also holds for convex mixtures of the w^D so by the Representation Theorem [33, Theorem 25.1]

$$w(R(a_1,a_2) \wedge R(a_1,a_3) \wedge R(a_4,a_2)) \geq w(R(a_1,a_2) \wedge R(a_3,a_4) \wedge R(a_5,a_6))$$

for any probability function w on $S\mathcal{L}$ satisfying Ex. Since by Ex

$$w(R(a_1,a_3) \wedge R(a_4,a_2)) = w(R(a_3,a_4) \wedge R(a_5,a_6))$$

the denominators of the conditional probabilities on both sides of the inequality in the statement of the theorem are the same and the required result now follows. □

Along similar lines, although much easier, we can show that (d) $\leq$ (e). For, when $T = \sum_{i,j} d_{i,j} = \sum_i A_i$ as before,

$$w^D(R(a_1,a_2) \wedge R(a_3,a_4) \wedge R(a_5,a_6)) = T^3 N^{-6}$$

$$w^D(R(a_1,a_2) \wedge R(a_1,a_4) \wedge R(a_5,a_6)) = TN^{-5}(\sum_i A_i^2)$$

and

$$T^3 N^{-6} \leq TN^{-5}(\sum_i A_i^2)$$

follows by Hölder's Inequality. Thus for w a probability function on $S\mathcal{L}$ satisfying Ex,

$$w(R(a_1,a_2) \wedge R(a_3,a_4) \wedge R(a_5,a_6)) \leq w(R(a_1,a_2) \wedge R(a_1,a_4) \wedge R(a_5,a_6)).$$

Since w satisfies Ex,

$$w(R(a_3,a_4) \wedge R(a_5,a_6)) = w(R(a_1,a_4) \wedge R(a_5,a_6))$$

and (d) $\leq$ (e) follows.

12

Turning to the other positive inequalities, (a) $\leq$ (b) is just Theorem 1 whilst (b) $\leq$ (d) and (c) $\leq$ (f) follow similarly by applying the Principle of Instantial Relevance, PIR, see [13], [22], [33] (or for the former using the footnote to Theorem 2). Furthermore, (b) $\leq$ (c) is (3) .

We now address the inequalities that do not (necessarily) hold. To give a counterexample to (c) $\leq$ (g) consider $w = w^D$ where D is an $N \times N$ $\{0, 1\}$-matrix. In this case, with the above abbreviations,

$$w(R(a_1, a_2) \,|\, R(a_1, a_4) \wedge R(a_5, a_2)) = T^{-2} \sum_{i,j} d_{i,j} A_i B_j, \tag{5}$$

$$w(R(a_1, a_2) \,|\, R(a_1, a_4)) = T^{-1} N^{-1} \sum_i A_i^2. \tag{6}$$

Letting D be the 6×6 matrix

$$
\begin{array}{cccccc}
1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
$$

gives values to (5), (6) of 9/32 and 7/24 respectively, and provides the required counterexample. This same probability function also gives a counterexample to (e) $\leq$ (g) (hence also to (e) $\leq$ (d)).

A counterexample to (d) $\leq$ (c) is given by setting

$$w\left(\bigwedge_{i,j\leq n} R(a_i, a_j)^{\epsilon_{i,j}}\right) = v\left(\bigwedge_{i,j\leq n} P(a_j)^{\epsilon_{i,j}}\right) \tag{7}$$

where v is any $c_\lambda^{L_1}$ of Carnap's Continuum of Inductive Methods[11] for the unary language $L_1 = \{P\}$ with $0 < \lambda < \infty$. In this case (d) $\leq$ (c) becomes

$$c_\lambda^{L_1}(P(a_2) \,|\, P(a_4) \wedge P(a_6)) \leq c_\lambda^{L_1}(P(a_2) \,|\, P(a_4)).$$

Putting in the actual values here gives

$$\frac{2 + 2^{-1}\lambda}{2 + \lambda} \leq \frac{1 + 2^{-1}\lambda}{1 + \lambda}$$

---

[11]See [3], or in the notation of this paper [33, Chapter 16].

which is false for λ in this range.

Similarly this probability function also gives counterexamples to (e) $\leq$ (c), (f) $\leq$ (c) and (b) $\leq$ (a).

We now describe a counterexample to (g) $\leq$ (f) . Let v be a probability function on the sentences of the language $\{P_1, P_2, P_3, \ldots\}$ where the P_i are unary, satisfying Px and Ex, and set

$$w\left(\bigwedge_{i,j\leq n} R(a_i, a_j)^{\epsilon_{i,j}}\right) = v\left(\bigwedge_{i,j\leq n} P_j(a_i)^{\epsilon_{i,j}}\right). \tag{8}$$

Then w satisfies Ex and the instance of (g) $\leq$ (f) for this w becomes

$$v(P_2(a_1) \,|\, P_4(a_1) \wedge P_6(a_1)) \geq v(P_2(a_1) \,|\, P_3(a_1) \wedge P_2(a_4))$$

and this can be seen to fail for the probability function v which treats the predicates as stochastically independent and identically distributed as, say, $c_2^{L_1}$ on $L_1 = \{P\}$. This same probability function furnishes a counterexample to (g) $\leq$ (e) and (g) $\leq$ (d).

To show that we do not have (c) $\leq$ (e) let D_1, D_2 be respectively the 4×4 matrices

$$
\begin{array}{cccc}
1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{cccc}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0
\end{array}
$$

Then,

$$
\begin{aligned}
w^{D_1}(R(a_1, a_4)) &= 7/16, \\
w^{D_1}(R(a_1, a_4) \wedge R(a_5, a_6)) &= (7/16)^2, \\
w^{D_1}(R(a_1, a_2) \wedge R(a_1, a_4)) &= (4^2 + 1^2 + 1^2 + 1^2)/4^3 = 19/64, \\
w^{D_1}(R(a_1, a_2) \wedge R(a_1, a_4) \wedge R(a_5, a_6)) &= (19/64) \times (7/16) = 133/1024, \\
w^{D_2}(R(a_1, a_4)) &= 8/16 = 1/2, \\
w^{D_2}(R(a_1, a_4) \wedge R(a_5, a_6)) &= (1/2)^2 = 1/4, \\
w^{D_2}(R(a_1, a_2) \wedge R(a_1, a_4)) &= (2^2 + 2^2 + 2^2 + 2^2)/4^3 = 1/4, \\
w^{D_2}(R(a_1, a_2) \wedge R(a_1, a_4) \wedge R(a_5, a_6)) &= (1/4 \times (1/2) = 1/8.
\end{aligned}
$$

14

Hence for $w = (w^{D_1} + w^{D_2})/2$,

$$w(R(a_1, a_2) \mid R(a_1, a_4)) = \frac{(19/64) + (1/4)}{(7/16) + (1/2)} = 7/12,$$

$$w(R(a_1, a_2) \mid R(a_1, a_4) \wedge R(a_5, a_6)) = \frac{(133/1024) + (1/8)}{(7/16)^2 + (1/4)} = 261/452,$$

giving a counterexample to (c) $\leq$ (e).

Finally to give an example where (d) $\leq$ (f) fails let E_1, E_2, respectively, be the 2×2 matrices

$$
\begin{array}{cc}
1 \quad 1 & \qquad 1 \quad 0 \\
1 \quad 1 & \qquad 0 \quad 0
\end{array}
$$

Then

$$
\begin{aligned}
w^{E_1}(R(a_1, a_4) \wedge R(a_1, a_6)) &= (2^2 + 2^2)/2^3 = 1, \\
w^{E_1}(R(a_1, a_2) \wedge R(a_1, a_4) \wedge R(a_1, a_6)) &= 1, \\
w^{E_1}(R(a_3, a_4) \wedge R(a_5, a_6)) = &= 1, \\
w^{E_1}(R(a_1, a_2) \wedge R(a_3, a_4) \wedge R(a_5, a_6)) &= 1, \\
w^{E_2}(R(a_1, a_4) \wedge R(a_1, a_6)) &= (1^2 + 0^2)/2^3 = 1/8, \\
w^{E_2}(R(a_1, a_2) \wedge R(a_1, a_4) \wedge R(a_1, a_6)) &= (1^3 + 0^3)/2^4 = 1/16, \\
w^{E_2}(R(a_3, a_4) \wedge R(a_5, a_6)) &= (1/4) \times (1/4) = 1/16, \\
w^{E_2}(R(a_1, a_2) \wedge R(a_3, a_4) \wedge R(a_5, a_6)) &= (1/64),
\end{aligned}
$$

Hence for $w = (w^{E_1} + w^{E_2})/2$,

$$w(R(a_1, a_2) \mid R(a_1, a_4) \wedge R(a_1, a_6)) = \frac{1 + (1/16)}{1 + (1/8)},$$

$$w(R(a_1, a_2) \mid R(a_3, a_4) \wedge R(a_5, a_6)) = \frac{1 + (1/64)}{1 + (1/16)},$$

giving a counterexample to (d) $\leq$ (f).

Combining the inequalities and non-inequalities produced above now gives exactly the relationships indicated in figure (i).

As already mentioned (a)-(g) do not form an exhaustive list of all such conditionals with at most two individually structurally similar sentences. Firstly each of these has a 'mirror image' formed by transposing the two coordinates, for example

(e) $w(R(a_1, a_2) \mid R(a_1, a_4) \wedge R(a_5, a_6))$,　(e') $w(R(a_2, a_1) \mid R(a_4, a_1) \wedge R(a_6, a_5))$.

15

Of these (a),(b),(d),(g) are the same as their mirror images while for the others it is easy to see by the device used with (7) or (8) that including them in figure (i) does not give any new vertical lines beyond those already present in the mirror image of figure (i).

More significantly we have left for future consideration the seemingly thorny cases of conditionals where the two supporting structurally similar sentences contain common constants which are not also common to the conditioned sentence, for example $w(R(a_1, a_2) \,|\, R(a_1, a_3) \wedge R(a_4, a_3))$.

Other Analogy Principles in PIL

In this short section we briefly mention some other analogy principles which have been suggested within the context of PIL as being in some sense rational. Ideally such principles will both capture a facet of analogical support as we intuit it and be consequences of other established and acknowledged rational principles (such as is the case with the Counterpart Principle) from which they too will inherit this status. Failing that they will present and formalize altogether new aspects of what might be understood as rational though their acceptability will then to a large degree hinge on their being consistent with other established and acknowledged rational principles.

Already in the joint paper [7] with Stegmüller Carnap was interested in the idea of analogical support within the context of his (unary) Inductive Logic and this was to continue right up to [6, Chapter 16]. The probability functions c_λ of Carnap's Continuum of Inductive Methods were too restrictive to encompass modeling the sort of analogical effects that Carnap had in mind which has led a number of investigators to propose instead various mixtures, products and generalizations of the c_λ, see for example [4], [6], [8], [11], [15], [23], [25], [26], [27], [28], [29], [36], [38], [39]. Such approaches however have so far still failed to fully achieve the required effects (see for example [24, page 81], [18], [19], [27], [35], [39], [40]). Furthermore being largely chosen to satisfy a property rather than being derived as the functions characterizing some concrete, arguably rational, principles they seem to us to suffer from a certain arbitrariness in the choices of the constituent functions, factors.

Two notable exceptions founded on general principles are di Maio's, [28], and Maher's [26]. Unfortunately neither seem entirely satisfactory, di Maio's requires a questionable (to our mind) linearity condition, [28, II3, page 378], and Maher's is effectively restricted to a language with at most two predicates (see [27]).

In [5] and [7] Carnap and Stegmüller also suggested a somewhat less arbitrary

notion of analogical support by similarity[12] based on sharing properties and a *derived* notion of closeness. To explain this let $R_1, R_2, \ldots, R_q$ be, as usual, the now all unary, predicates of L and let $\alpha_1(x), \alpha_2(x), \ldots, \alpha_{2^q}(x)$ be the atoms of L, that is the formulae of the form

$$R_1^{\epsilon_1}(x) \wedge R_2^{\epsilon_2}(x) \wedge \ldots \wedge R_q^{\epsilon_q}(x)$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_q \in \{0, 1\}$. Then for atoms

$$\alpha_r(x) = \bigwedge_{i=1}^{q} R_i^{\epsilon_i}(x), \quad \alpha_m(x) = \bigwedge_{i=1}^{q} R_i^{\delta_i}(x),$$

the analogical support afforded to $\alpha_m(a_2)$ by $\alpha_r(a_1)$, in the absence of any further information on a_1, a_2, was to be a decreasing function (in terms of the subset ordering) of the set $\Delta(\alpha_r, \alpha_m) = \{i \mid \epsilon_i = \delta_i\}$.

More specifically Carnap's writings suggest an analogy principle (see [5, page 320], [17, page 101]) of the form:

For atoms $\alpha_r, \alpha_m, \alpha_k$, if $\Delta(\alpha_r, \alpha_m) \subseteq \Delta(\alpha_r, \alpha_k)$ then

$$w(\alpha_r(a_{n+2}) \mid \alpha_m(a_{n+1}) \wedge \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)) \leq w(\alpha_r(a_{n+2}) \mid \alpha_k(a_{n+1}) \wedge \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)).^{[13]} \quad (9)$$

Using Johnson's Sufficientness Postulate it is easy to see that this principle holds for the probability functions c_λ in Carnap's Continuum of Inductive Methods. Unfortunately this is somewhat less satisfactory than it might initially appear because for the c_λ^L (9) holds with equality whenever $m, k \neq r$,[14] hardly capturing the idea that $\alpha_m(a_{n+1})$'s support for $\alpha_r(a_{n+2})$ *depends* on the set of shared features. Several families of probability functions are known which do satisfy (9) with strict subset and inequality but currently no complete characterization is available (see for example [10, Section 3], [15], [26], [40]).

A second point on which (9) might be queried is the requirement that the standing evidence should be of the form $\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)$, in other words a state description,

---

[12]In [6] Carnap also has a notion of analogy by proximity but that essentially denies Ex so is not under consideration here.

[13]In the accounts related in this section the problem of conditional probabilities being undefined when the denominator probability is zero was avoided by adopting the convention of treating $w(\theta \mid \phi) \geq w(\theta \mid \psi)$ (etc.) as an abbreviation for $w(\theta \wedge \phi)w(\psi) \geq w(\theta \wedge \psi)w(\phi)$.

[14]In fact with Atom Exchangeability and Regularity this characterizes the c_λ^L for $0 < \lambda \leq \infty$, see [9].

rather than, say, simply a sentence $\phi(a_1, a_2, \ldots, a_n) \in QFSL$. After all the underlying intuition in both cases would seem to be no different. In fact allowing this generalization has a significant effect. With the exception of c_0^L, c_∞^L the principle no longer holds for Carnap's Continuum even when $q = 2$. Indeed there are only a very few further probability functions satisfying the principle (and Ex+Px+SN[15]), see [19] for a complete characterization, none of them seemingly particularly attractive as a rational choice.

More recently several other analogy principles based on interpretations of the earlier mentioned Candidate Analogical Inference Rule (R) of Bartha, [1], have been investigated in the context of PIL, see [21], in particular a version of this rule which prescribes that the more properties on which a target and source agree (and the fewer on which they disagree) the more support there is that the target will satisfy an additional property known to be satisfied by the source. While this allows considerable freedom of formalization within PIL the basic version seems to be the principle:

For $\theta(a_{n+1}, \vec{a}), \phi(a_{n+1}, \vec{a}) \in QFSL$, where $\vec{a} = \langle a_1, a_2, \ldots, a_n \rangle$,

$$w(\phi(a_{n+2}, \vec{a}) \,|\, \theta(a_{n+1}, \vec{a}) \wedge \theta(a_{n+2}, \vec{a}) \wedge \phi(a_{n+1}, \vec{a})) \geq w(\phi(a_{n+2}, \vec{a}) \,|\, \phi(a_{n+1}, \vec{a})).$$

That is the similarity of a_{n+1}, a_{n+2} engendered by them both satisfying θ argues that a_{n+2} will satisfy ϕ given that a_{n+1} does.

Unfortunately c_0^L is the only probability function satisfying this principle (together with Ex+Px+SN). In [21] various refinements of this principle were considered some of which do follow from commonly accepted rational principles within PIL. Nevertheless the failure of the above basic version must put a question mark against this formalization of the underlying intuition grounding rule (R), at least within the framework of PIL.

Conclusion

In this paper we have considered all possible variations of the Counterpart Principle of Analogical Support by Structural Similarity which hold under the assumption of Language Invariance, Li, when we allow up to two analogies and all relation/constant symbols common to the conditioning conjuncts are also present in the conditioned sentence.[16] In summary these turn out to be generated by the

---

[15]SN is the *Strong Negation Principle* which asserts that if θ' is the result of replacing a relation symbol R everywhere in θ by $\neg R$ then $w(\theta') = w(\theta)$.

[16]In this paper the relationship of, say, $w(R(a_1, a_2) \,|\, R(a_1, a_3) \wedge R(a_4, a_3))$ to (a)-(g) is not considered.

template of particular instances

$$
\begin{array}{lll}
\text{(b)} \geq \text{(a):} & w(R(a_1, a_2) \,|\, R(a_3, a_4)) & \geq & w(R(a_1, a_2)) \\
\text{(c)} \geq \text{(b):} & w(R(a_1, a_2) \,|\, R(a_1, a_4)) & \geq & w(R(a_1, a_2) \,|\, R(a_3, a_4)) \\
\text{(d)} \geq \text{(b):} \; w(R(a_1, a_2) \,|\, R(a_3, a_4) \wedge R(a_5, a_6)) & \geq & w(R(a_1, a_2) \,|\, R(a_3, a_4)) \\
\text{(e)} \geq \text{(d):} \; w(R(a_1, a_2) \,|\, R(a_1, a_4) \wedge R(a_5, a_6)) & \geq & w(R(a_1, a_2) \,|\, R(a_3, a_4) \wedge R(a_5, a_6)) \\
\text{(g)} \geq \text{(d):} \; w(R(a_1, a_2) \,|\, R(a_1, a_4) \wedge R(a_5, a_2)) & \geq & w(R(a_1, a_2) \,|\, R(a_3, a_4) \wedge R(a_5, a_6)) \\
\text{(f)} \geq \text{(c):} \; w(R(a_1, a_2) \,|\, R(a_1, a_4) \wedge R(a_1, a_6)) & \geq & w(R(a_1, a_2) \,|\, R(a_1, a_4))
\end{array}
$$

for w a probability function on the language $\{R\}$ satisfying Ex. By a template we mean that $R(a_i, a_j)$ may be replaced here by a sentence of a language L in which the a_i substitute for disjoint blocks of constant and/or relation symbols of L and w is replaced by a probability function on L satisfying Li.

From (b)≥(a), (d)≥(b) and (f)≥(c) we see that, as one might expect, more analogies of the same structural similarity produce more support (or at least not strictly less support) whilst from (c)≥(b), (e)≥(d) and (g)≥(d) we have, again not unexpectedly, that greater structural similarity gives greater analogical support. More surprising are some of the inequalities which do not necessarily hold, for example that we need not have (e)≥(c). Here adding to the conditioning $R(a_1, a_4)$ in (c) the analogous but more distantly similar $R(a_5, a_6)$ can actually reduce the analogical support, though support it remains since from (e)≥(b), (b)≥(a) we do still have (e)≥(a). In terms of our introductory example this says that although adding the information about the aunt and *The Sound of Music* still leaves some analogical support, it may in fact diminish the support provided by the *Toy Story* evidence alone.

The results in this paper are clearly just the tip of the iceberg, for example we have not considered the case where the underlying template involves not simply binary R but ternary, 4-ary etc. Even in the case of binary R we have not considered cases where constants common to the conditioning sentences are not also present in the conditioned sentence. We have also not considered cases where we also have negated analogies (as allowed by Bartha in his 'Candidate Analogical Inference Rule' described in [21]). Given our results so far, and the technicalities needed to show them, the challenge of answering the apparently limitless number of genuinely new questions that arise, and in turn of building those answers into our understanding, seems formidable. Certainly at this juncture the prospect of an elegant, meaningful *general theory of multiple analogical support by structural similarity* can be no more than a matter of pure speculation.

Finally it is worth reminding ourselves that the results in this paper have been

derived on the assumption that one's chosen rational probability function satisfies just Li. In Polyadic Inductive Logic however there are a number of stronger 'rationality principles' which have been proposed, in particular *Li with Spectrum Exchangeability*,[17], and it may be that making these further assumptions will produce a different picture. Given our previous difficulties in attempting to generalize PIR to this polyadic context (see [33, Chapter 36]) that too would look to be a challenging endeavour.[18]

References

[1] Bartha, P., Analogy and Analogical Reasoning, *The Stanford Encyclopedia of Philosophy (Fall 2013 Edition)*, ed. E.N.Zalta, URL = http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy/

[2] Bartha, P., *By Parallel Reasoning,* Oxford University Press, 2009.

[3] Carnap, R., *The Continuum of Inductive Methods,* University of Chicago Press, 1952.

[4] Carnap, R., $m(Z)$ *for n Families as a Combination of* m_λ*-Functions*, document 093-22-01 (see also 093-22-03) in the Rudolf Carnap Collection, University of Pittsburg Library, 1954.

[5] Carnap, R., Notes on Probability and Induction, in *Logical Empiricist*, ed. J.Hintikka, Reidel, Dordrecht, 1973, pp. 293-324.

[6] Carnap, R., A Basic System of Inductive Logic Part II, in *Studies in Inductive Logic and Probability, Volume II*, ed. R.C.Jeffrey, University of California Press, 1980, pp. 7-155.

[7] Carnap, R. & Stegmüller, W., *Induktive Logik und Wahrscheinlichkeit,* Springer Verlag, Wien, 1959.

[8] Costantini, D., Analogy by Similarity, *Erkenntnis*, 1983, **20**, 107-114.

---

[17]In a nutshell Spectrum Exchangeability asserts that the probability of a state description should depend only on the numbers of constants which look identical according to that state description and not on what particular relations they satisfy, (see for example [33]). It is a natural extension to polyadic languages of Atom Exchangeability, so essentially Carnap's Attribute Symmetry, [6, page 77], for unary languages.

[18]Another approach to PIR which may also shed some light on analogy can be found in [37].

[9] D'Asaro, F.A. & Paris, J.B., A Note on Carnap's Continuum and the Weak State Description Analogy Principle. Available at www.maths.manchester.ac.uk/∼jeff/papers/fad160208sdah

[10] D'Asaro, F.A., *Analogical Reasoning in Unary Inductive Logic,* Master's Thesis, University of Manchester, 2014. Available at http//www.maths.manchester.ac.uk/∼jeff/theses/fdathesis.pdf

[11] Festa, R., Analogy and Exchangeability in Predictive Inferences, *Erkenntnis,* 1996, **45**, 229-252.

[12] Gaifman, H., Concerning Measures on First Order Calculi, *Israel Journal of Mathematics*, 1964, **2**, 1-18.

[13] 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, 1971, pp. 235-251.

[14] Gentner, D., Structure-Mapping: A Theoretical Framework for Analogy, *Cognitive Science,* 1983, **7**, 155-170.

[15] Hesse, M., Analogy and Confirmation Theory, *Philosophy of Scince*, 1964, **31**, 319-327.

[16] Hesse, M., *Models and Analogies in Science,* University of Notre Dame Press, 1966.

[17] Hill, A., *Reasoning by Analogy in Inductive Logic*, Doctorial Thesis, University of Manchester, 2013. Available at http://www.maths.manchester.ac.uk/∼jeff/theses/ahthesis.pdf

[18] Hill, A. & Paris, J.B., Reasoning by Analogy in Inductive Logic, *The Logica Yearbook 2011,* eds. M.Peliš & V.Punčochář, College Publications, London, 2012.

[19] Hill, A. & Paris, J.B., An Analogy Principle in Inductive Logic, *Annals of Pure and Applied Logic,* 2013, **164**, 1293-1321.

[20] Hill, A. & Paris, J.B., The Counterpart Principle of Analogical Support by Structural Similarity, *Erkenntnis*, 2013, **79**(2), 1-16.

[21] Howarth, E., Paris, J.B. & Vencovská, A., An examination of the SEP Candidate Analogical Inference Rule within Pure Inductive Logic, *Journal of Applied Logic,* 2016, **14**:22-45.

21

[22] Humburg, J., The Principle of Instantial Relevance, *Studies in Inductive Logic and Probability*, eds. R.Carnap & R.C.Jeffrey, vol.I, University of California Press, Berkeley and Los Angeles, 1971, pp. 225-233.

[23] Huttegger, S.M., Analogical Predictive Probabilities, 2014, available at http://faculty.sites.uci.edu/shuttegg/files/2011/03/AnalogicalPredictionV4.pdf

[24] Kuipers, T.A.F., *From Instrumentalism to Constructive Realism,* Synthese Library/Volume 287, 2000.

[25] Kuipers, T.A.F., Two types of Inductive Analogy by Similarity, *Erkenntnis*, 1984, **21**, 63-87.

[26] Maher, P., Probabilites 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] di Maio, M.C., Predictive Probability and Analogy by Similarity, *Erkenntnis*, 1995, **43**(3), 369-394.

[29] Niiniluoto, I., Analogy and Inductive Logic, *Erkenntnis*, **16**, 1981, pp. 1-34.

[30] Niiniluoto, I., 1988, Analogy and Similarity in Scientific Reasoning, in *Analogical Reasoning,* ed. D.H.Helman, Kluwer Academic Publishers, 1988, pp. 271-298.

[31] Paris, J.B., *Pure Inductive Logic: Workshop Notes for Progic 2015.* Available at http://www.maths.manchester.ac.uk/∼jeff/lecture-notes/Progic15.pdf

[32] Paris, J.B., *An Equivalent of Language Invariance.* Available at MIMS EPrints, http://eprints.ma.man.ac.uk/2437/

[33] Paris, J.B. & Vencovská, A., *Pure Inductive Logic*, in the Association of Symbolic Logic Perspectives in Mathematical Logic Series, Cambridge University Press, April 2015.

[34] Paris, J.B. & Vencovská, A., *An inequality concerning the expected values of row-sum and column-sum products in Boolean matrices.* Available at MIMS EPrints, http://eprints.ma.man.ac.uk/2469/

[35] Pietarinen, J., *Lawlikeness, Analogy and Inductive Logic,* Amsterdam, North-Holland Pub. Co., 1972.

[36] Romeijn, J.W., Analogical Predictions for Explicit Similarity, *Erkenntnis,* 2006, **64**(2), 253-280.

[37] Ronel, T. & Vencovská, A., The Principle of Signature Exchangeability, *Journal of Applied Logic,* May 2016, **15**:16-45.

[38] Skyrms, B., Analogy by Similarity in Hyper-Carnapian Inductive Logic, *Philosophical Problems of the Internal and External Worlds,* eds. J.Earman, A.I.Janis, G.J.Massey & N.Rescher, University of Pittsburg Press/Universitätsverlag Konstanz, 1993, pp. 273-282.

[39] Spohn, W., Analogy and Inductive Logic: A Note on Niiniluoto, *Erkenntnis*, 1981, **16**, 35-52.

[40] Welch, J., Singular Analogy and Quantitative Inductive Logics, *Theoria*, 1999, **14**(2), 207-247.