Pure Inductive Logic,
Workshop Notes for Progic 2015

Jeff Paris
School of Mathematics
The University of Manchester
Manchester M13 9PL

jeff.paris@manchester.ac.uk

April 9, 2015

Introduction Broadly speaking Inductive Logic is the study of how to rationally or logically assign probabilities, subjective probabilities, to events on the basis of some knowledge. *Pure Inductive Logic* is intended to address this question in perhaps the simplest possible case, when there actually is no prior contingent knowledge, just the uninterpreted context.

Of course we cannot expect such a study to immediately provide practical answers but nevertheless it might well give us some insight into the more complicated situations the real world presents to us – and in any case if we cannot provide answers in this very simplified case whyever should we expect to do better when things become more complicated?

In the tradition of the subject as originally presented by W.E.Johnson and R.Carnap in the 1920's-40's (see [2], [3], [4], [16]) we shall initially take this basic, simple, context to be a first order structure M for the language L_q with just countably many constant symbols a_n, $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$ and predicate (i.e. unary relation) symbols R_i, $i = 1, 2, \ldots, q$ and such that the interpretation of the a_n in M (which we shall also denote a_n) lists all the elements of the universe of M (though not necessarily without repeats). In particular for these notes we will not allow function symbols nor equality in the language L. For future reference let $SL_q, QFSL_q$ denote the sentences, quantifier free sentences of L_q.

The question we are interested in can be thought of as follows:

> Q: Given an agent $\mathcal{A}$ inhabiting M and $\theta \in SL_q$, rationally or logically, what probability should $\mathcal{A}$ assign to θ?

More fully, since we can reasonably demand that to be rational these probabilities for different $\theta \in SL$ should be coherent we are really asking what *rational*

probability function w should $\mathcal{A}$ adopt? where $w : SL_q \to [0,1]$ is a *probability function on* L_q if it satisfies that for all $\theta, \phi, \exists x\,\psi(x) \in SL_q$

(P1) $\models \theta \;\Rightarrow\; w(\theta) = 1.$

(P2) $\theta \models \neg\phi \;\Rightarrow\; w(\theta \vee \phi) = w(\theta) + w(\phi).$

(P3) $w(\exists x\,\psi(x)) = \lim_{n\to\infty} w(\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n)).$

Condition (P3) is often referred to as *Gaifman's Condition*, see [8], and is a special addition to the conventional conditions (P1), (P2) appropriate to this context. It intends to capture the idea that the $a_1, a_2, a_3, \ldots$ exhaust the universe.

All the standard, simple, properties you'd expect of a probability function follow from these (P1-3):

Proposition 1. *Let w be a probability function on SL. Then for $\theta, \phi \in SL$,*

(a) $w(\neg\theta) = 1 - w(\theta).$

(b) $\models \neg\theta \;\Rightarrow\; w(\theta) = 0.$

(c) $\theta \models \phi \;\Rightarrow\; w(\theta) \le w(\phi).$

(d) $\theta \equiv \phi \;\Rightarrow\; w(\theta) = w(\phi).$

(e) $w(\theta \vee \phi) = w(\theta) + w(\phi) - w(\theta \wedge \phi).$

Proofs may be found in [24] or [26]. As usual for w a probability function on L_q we take the corresponding conditional probability $w(\theta \,|\, \phi)$ to be a probability function such that

$$w(\theta \,|\, \phi) \cdot w(\phi) = w(\theta \wedge \phi), \quad \text{i.e. } w(\theta \,|\, \phi) = \frac{w(\theta \wedge \phi)}{w(\phi)} \;\; \text{if } w(\phi) > 0.$$

With this in mind the question becomes:

> Q: Given an agent $\mathcal{A}$ inhabiting M, rationally or logically, what probability function w should $\mathcal{A}$ adopt?

It should be emphasized here that otherwise $\mathcal{A}$ knows nothing about M, s/he has no particular interpretation in mind for the constants and predicates.

On the face of it it might appear that because of the great diversity of sentences in SL_q probability functions would be very complicated objects and not easily described. In fact this is not the case as we shall now explain. The first step in this direction is the following theorem of Gaifman, [8]:[1]

---

[1] For a proof in the notation of these tutorials see Theorem 7 of [24].

Theorem 2. *Suppose that $w : QFSL_q \to [0,1]$ satisfies (P1) and (P2) for $\theta, \phi \in QFSL_q$. Then w has a unique extension to a probability function on SL_q satisfying (P1),(P2),(P3) for any $\theta, \phi, \exists x\, \psi(x) \in SL_q$.*

To give an example of a very simple probability function let

$$\vec{c} \in \mathbb{D}_{2^q} = \{\langle x_1, x_2, \ldots, x_{2^q}\rangle \mid x_i \geq 0,\ \sum_{i=1}^{2^q} x_i = 1\}$$

and define $w_{\vec{c}}$ for the *atoms* $\alpha_1, \ldots, \alpha_{2^q}$, that is for the 2^q formulae of the form

$$\pm R_1(x) \wedge \pm R_2(x) \wedge \ldots \wedge \pm R_q(x),$$

by

$$w_{\vec{c}}(\alpha_j(a_i)) = c_j, \quad j = 1, 2, \ldots, 2^q.$$

Notice that knowing which atom a_i satisfies already tells us everything there is to know about a_i per se.

Extend $w_{\vec{c}}$ to *state descriptions*, that is conjunctions of atoms, by setting, e.g.,

$$
\begin{aligned}
w_{\vec{c}}(\alpha_{h_1}(a_1) \wedge \alpha_{h_2}(a_2) \wedge \ldots \wedge \alpha_{h_n}(a_n)) &= w_{\vec{c}}(\alpha_{h_1}(a_1)) \times w_{\vec{c}}(\alpha_{h_2}(a_2)) \times \ldots \times w_{\vec{c}}(\alpha_{h_n}(a_n))) \\
&= c_{h_1} \times c_{h_2} \times \ldots \times c_{h_n} \\
&= \prod_{j=1}^{n} c_{h_j} = \prod_{k=1}^{2^q} c_k^{m_k} \quad \text{where } m_k = |\{j \mid h_j = k\}|.
\end{aligned}
$$

By the Disjunctive Normal Form Theorem any $\theta \in QFSL_q$ is logically equivalent to a disjunction of state descriptions, say,

$$\theta(a_1, a_2, \ldots, a_n) \equiv \bigvee_{k=1}^{s} \bigwedge_{i=1}^{n} \alpha_{h_{ik}}(a_i)$$

and we can extend $w_{\vec{c}}$ to $QFSL_q$ by setting

$$
\begin{aligned}
w_{\vec{c}}(\theta) &= w_{\vec{c}}\left(\bigvee_{k=1}^{s} \bigwedge_{i=1}^{n} \alpha_{h_{ik}}(a_i)\right) \\
&= \sum_{k=1}^{s} w_{\vec{c}}\left(\bigwedge_{i=1}^{n} \alpha_{h_{ik}}(a_i)\right) \\
&= \sum_{k=1}^{s} \prod_{i=1}^{n} w_{\vec{c}}(\alpha_{h_{ik}}(a_i)) \\
&= \sum_{k=1}^{s} \prod_{i=1}^{n} c_{h_{ik}}.
\end{aligned}
$$

3

From this example it should be clear that to specify a probability function it is enough to specify its values simply on state descriptions.

The key word here in $\mathcal{Q}$ is 'rational', but what does that mean? Well, at the present time we shall leave that to the reader's intuition, we all seem to have an appreciation of what we mean by rational, at least we certainly seem to recognize irrational behavior when we see it! One useful way to make this notion more concrete is to imagine that next door is another agent in the same situation as $\mathcal{A}$. These agents know nothing more of each other, nor can they communicate but nevertheless they are endeavoring to assign the *same* probabilities. In this case we might equate 'rationally assign' with this goal of conformity.

The method of answering this question $\mathcal{Q}$ within Pure Inductive Logic has been (up to now) to propose principle of probability assignment which it would arguably be irrational to flout and consider the restrictions they impose on the agent $\mathcal{A}$'s choice, and how these principles relate to each other. Ideally one might hope there are such principles which are both widely accepted and whose net effect is to determine $\mathcal{A}$'s possible choices uniquely. In that case we might feel justified in asserting that these principles *define* what we mean by 'rational' in this context.[2] Unfortunately no such consensus seems to be currently within sight, we have a number of principles which are ostensibly 'rational' to some extent and rather than reinforcing each other they are sometimes even disjoint.

To date there have been three main sources of such principles:

- **Symmetry**: The idea that $\mathcal{A}$'s choice should not break existing symmetries. For example in the absence of any additional information it would be irrational to give the probability of a coin toss coming down heads anything but one half since to do so would introduce an asymmetry between heads and tails which was not there initially.

- **Irrelevance**: The idea that irrelevant information should have no effect. For example the probability I would give to England winning the next cricket series against India should not be influence by learning that it rained today in Rome.

- **Relevance**: The idea that acquiring supporting evidence for a proposition should cause one to increase one's appraisal of its probability. For example hearing that India are 2 for 4 in the first innings against England should cause me to shorten the odds I'd give for an England victory.

Recently however we (Alex Hill, Liz Howarth, Alena Vencovská and myself) have been looking at another source of such principles, *Analogy*, and this what I'd like to say something about in these lectures. Before doing so however we need to consider some symmetry principles which are so well accepted as to be simply implicit basic assumptions in this subject.

---

[2]As we would argue happens in the propositional case, see [25].

The Constant Exchangeability Principle, Ex[3]

For $\phi(a_1, a_2, \ldots, a_m) \in SL_q$[4] and (distinct) $i_1, i_2, \ldots, i_m \in \mathbb{N}^+$,

$$w(\phi(a_1, a_2, \ldots, a_m)) = w(\phi(a_{i_1}, a_{i_2}, \ldots, a_{i_m})).$$

The rational justification here is that the agent has no knowledge about any of the a_i so it would be irrational to treat them differently when assigning probabilities.[5],

Constant Exchangeability is so widely accepted in this area that we will henceforth *assume it throughout without further explicit mention.*

The probability functions $w_{\vec{c}}$ satisfy Ex. Indeed they are *exactly* the probability functions satisfying Ex and the

The Constant Irrelevance Principle

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

$$w(\theta \wedge \phi) = w(\theta) \cdot w(\phi),$$

equivalently

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

For a proof of this see [26, Chapter 8]. The appellation 'Irrelevance' here obvious, the fact that θ and ϕ only give information about completely disjoint sets of constants is taken to mean that the probability of θ should not change if one was to learn ϕ. [The fact that this requirement restricts w to such a small, and, with the possible exclusion of $w_{\langle 2^{-q}, 2^{-q}, \ldots, 2^{-q} \rangle}$, hardly rational set of probability functions suggests that this principle possibly takes the idea of irrelevance too far.]

Now that we have introduced this Constant Exchangeability Principle, Ex, we would like to investigate what it entails, what follows from the assumption that w satisfies Ex. Often a major step in PIL after one has formulated a principle is to prove a representation theorem for the probability functions satisfying that principle by showing that they must look like a combination of certain 'simple building block functions'. There are such results for probability functions satisfying Ex, the first of these, and historically the most important, being the so call *de Finetti's Representation Theorem* for the case of the unary language L_q.

---

[3]Johnson's *Permutation Postulate* and Carnap's *Axiom of Symmetry.*

[4]The convention is that when a sentence ϕ is written in this form it is assumed (unless otherwise stated) that the displayed constants are distinct and include all the constants actually occurring in ϕ.

[5]The agent is not supposed to 'know' that a_1 comes before a_2 which comes before . . . . . in our list.

The following theorem due to Bruno de Finetti may be found in [7] (for a proof of this result in the notation being used here see [24, Theorem 10] or [26, Chapter 9]).

De Finetti's Representation Theorem 3. *A probability function w on a unary language L_q satisfies Ex just if it is a mixture of the $w_{\vec{x}}$.*

More precisely, just if

$$w = \int w_{\vec{x}} \, d\mu(\vec{x})$$

for some countably additive measure μ on the Borel subsets of

$$\{\langle x_1, x_2, \ldots, x_{2^q} \rangle \, | \, 0 \leq x_1, x_2, \ldots, x_{2^q}, \sum_i x_i = 1\}. \tag{1}$$

On the face of it this theorem may seem to only be of interest to mathematicians, it doesn't seem to be saying much about induction or rationality which would be of interest to a philosopher. However it yields consequences and observations which surely are of interest in this regard. The mathematical power of this theorem lies in the fact that it often enables us to translate questions about the general probability function w on the left hand side of (1) into questions about the very simple probability functions $w_{\vec{x}}$ on the right hand side. For example by this simple device Humburg [14] showed the following result, originally due to Gaifman [9]:

Theorem 4. *Ex implies the:*

(Extended) Principle of Instantial Relevance, PIR

For $\theta(a_1, a_2, \ldots, a_n), \phi(a_{n+1}) \in SL_q$,

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

Proof. Let the probability function w on L satisfy Ex. Write just θ for $\theta(a_1, a_2, \ldots, a_n)$.

Then for μ the de Finetti prior for w,

$$w(\theta) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta) \, d\mu(\vec{x}) = A \text{ say,}$$

$$w(\phi(a_{n+1}) \wedge \theta) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\phi(a_{n+1}) \wedge \theta) \, d\mu(\vec{x}) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\phi(a_{n+1})) \cdot w_{\vec{x}}(\theta) \, d\mu(\vec{x}) = B \text{ say,}$$

$$w(\phi(a_{n+2}) \wedge \phi(a_{n+1}) \wedge \theta) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\phi(a_{n+1}))^2 \cdot w_{\vec{x}}(\theta) \, d\mu(\vec{x}),$$

and (2) amounts to

$$\left(\int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\phi(a_{n+1})) \cdot w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \right)^2 \le$$

$$\left(\int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \right) \cdot \left(\int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\phi(a_{n+1}))^2 \cdot w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \right) \qquad (3)$$

If $A = 0$ then this clearly holds, so assume that $A \ne 0$.

Then multiplying out

$$
\begin{aligned}
0 \;\le\; & \int_{\mathbb{D}_{2^q}} \left(w_{\vec{x}}(\phi(a_{n+1}))A - \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\phi(a_{n+1}) \cdot w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \right)^2 w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \\
=\; & \int_{\mathbb{D}_{2^q}} \left(A w_{\vec{x}}(\phi(a_{n+1})) - B \right)^2 w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \\
=\; & \int_{\mathbb{D}_{2^q}} \left(A^2 w_{\vec{x}}(\phi(a_{n+1}))^2 - 2AB w_{\vec{x}}(\phi(a_{n+1})) + B^2 \right) w_{\vec{x}}(\theta) \, d\mu(\vec{x}) \\
=\; & \int_{\mathbb{D}_{2^q}} A^2 w_{\vec{x}}(\phi(a_{n+1}))^2 w_{\vec{x}}(\theta) \, d\mu(\vec{x}) - 2AB^2 + AB^2
\end{aligned}
$$

and dividing by A gives (3), and the result follows. $\qquad\square$

From a philosopher's point of view this *is* an interesting result (or at least it should be!) because it confirms one's intuition re induction that the more often you've seen something in the past the more probable you should expect it to be in the future. So this turns out to be simply a consequence of Ex. Whilst one might claim, in accord with Hume [15], that Ex is really the assumption of the uniformity of nature and in that sense can be equated with induction[6] what this argument does show is that *if you think Ex rational then you should think PIR rational.*

It is also worth emphasizing here that we have this *relevance* principle is actually a consequence of the *symmetry* principle Ex. This appears to be a common feature of PIL, that recognizably rational manifestations of relevance come simply as derivatives of symmetry rather than fundamental notions in their own right.

In an exactly similar way to the Constant Exchangeability Principle Ex we can justify another symmetry principle, the

Principle of Predicate Exchangeability, Px

For $\phi(P_1, P_2, \ldots, P_m) \in SL_q$, where we explicitly display the predicate symbols occurring in ϕ, and (distinct) $1 \le i_1, i_2, \ldots, i_m \le q$,

$$w(\phi(P_1, P_2, \ldots, P_m)) = w(\phi(P_{i_1}, P_{i_2}, \ldots, P_{i_m})).$$

---

[6]The result does not go the other way however, see [29, footnote 6] for an example of a probability function satisfying PIR but not Ex.

Px differs from Ex in the sense that in L_q there are only finitely many predicates to permute whereas there are infinitely many constants. However this will in effect be remedied once we introduce our next rationality requirement.

Consider again our agent $\mathcal{A}$ inhabiting a structure for L_q and suppose that $\mathcal{A}$ *imagined* that the language also contained predicate symbols in addition to those in L_q. In that case would not $\mathcal{A}$ then wish to adopt a probability function for that larger language which satisfied the same principle that s/he would have considered rational for L_q alone? Surely yes!

But then can this wish of $\mathcal{A}$ actually be realized? The problem is that $\mathcal{A}$ might follow his/her rational principles and pick the probability functions w on L_q and w^+ on the (imagined) extended language L^+ and find that the restriction of w^+ to SL, denoted $w^+ \restriction SL_q$, is not the same as w. In other words simply by imagining being rational in L^+ the agent would have discredited w. Indeed looked at from this perspective w might seem a particularly bad choice if there was no extension at all of w to L^+ which satisfied the agent's favored rational principles.

To make this more concrete suppose the agent felt Ex + Px was the (only) rational requirement that s/he was obliged to impose on his/her choice w. Then it might be that the agent made such a choice only to realize that there was no way to extend this probability function to a larger language and still maintain having Ex + Px.

In fact this can happen for some bad choices of w, but fortunately it needn't happen, there will be choices of probability function for which there are no such dead ends. These are the ones which satisfy:

Unary Language Invariance, ULi

A probability function w satisfies Unary Language Invariance if there is a family of probability functions w^r, one on each language L_r, containing w (so $w = w^q$) such that each member of this family satisfies Px and whenever $p \leq r$ then $w^r \restriction SL_p = w^p$.

The $w_{\vec{c}}$ do not satisfy ULi, or even Px, except for rather special choices[7] of $\vec{c}$. However many of the key (as far as PIL is concerned) mixtures of them do, for example Carnap's Continuum of Inductive Methods, see [1], and the Nix-Paris Continuum, see [23] or [26].

To simplify some later results it will be useful to introduce one more symmetry principle:

---

[7]For ULi the requirement is that

$$c_i = \int_{[0,1]} y^j (1-y)^{q-j} \, d\mu(y)$$

for some measure μ on $[0,1]$, where j is the number of negations appearing in α_i, see [18].

The Strong Negation Principle, SN

For $\theta \in SL_q$,

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

where θ' is the result of replacing each occurrence of the relation symbol R in θ by $\neg R$.

SN is essentially what is being used when we argue for rationality of the heads/tails symmetry when we toss a coin, since not-heads is just tails. Viewed in this way then SN seems, like Ex and Px, evidently rational.

We now turn our attention to the main goal of this tutorial, reasoning by analogy, or more precisely the investigation of 'rational principles of analogical support'. Reasoning by analogy is often asserted to be *the* central method in many traditional Eastern Logics, in particular the Navya-Nyaya logic of India[8] and the Late Mohist logic of China. And clearly it is also extremely widely adopted 'on the street', much more common in fact than ever applying the rules of the Propositional or Predicate Calculi. For example if I know that my 10 year old nephew enjoyed a Christmas party in the local swimming pool I might recommend this possibility to a colleague who wishes to arrange a birthday party for her 9 year old daughter.

Because of its ubiquity attempting to formalize what is meant by 'analogical support', and in turn why it is rational or logical, has, over the years, received much attention from philosophers in the context of Inductive Logic, for example [3], [4], [5] [6], [17], [19], [20], [21], [22], [27], [28]. A common approach is to relate the analogical support that $\alpha_i(a_1)$ gives to $\alpha_j(a_2)$ to the *Hamming distance* $|\alpha_i - \alpha_j|$ of the atom $\alpha_i(x)$ from $\alpha_j(x)$, where for $\alpha_i(x) = \bigwedge_{n=1}^{q} R_n^{\epsilon_n}(x)$, $\alpha_j(x) = \bigwedge_{n=1}^{q} R_n^{\delta_n}(x)$,

$$|\alpha_i - \alpha_j| = \sum_{n=1}^{q} |\epsilon_n - \delta_n|,$$

the number of conjuncts R_n on which the atoms differ. In other words the $\pm R_n$ which α_j has in common with α_i provide support for α_i and those which they do not have in common oppose α_i. This idea leads to the

Principle of Analogical Support by Distance:

If $\theta(a_1, a_2, \ldots, a_n) \in QFSL_q$ and

$$|\alpha_i - \alpha_j| < |\alpha_i - \alpha_k|$$

then

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

Unfortunately, as is shown in [11], even in the case of a language with just two predicates i.e. L_2, very few probability functions satisfy this principle together

---

[8]Which is precisely why I'm focusing on analogy in these tutorials!

with the Principle of Strong Negation[9] and our standing assumptions Ex and Px
and seemingly those are not ones that have much of a claim to being rational.
The choice becomes less still if we increase two predicates to three and indeed
there are then none at all if we make the inequality in the conclusion of this
principle strict. [See [10] for alternative, but no less disheartening, combinations
of initial symmetry assumptions.] Some hope of salvation here does currently
seem to exist if we restrict $\theta \in QFSL_q$ here to being a state description, it is
not clear that version is so hard to fulfill (see [11]), however this does seem a
rather artificial restriction to place on θ in the circumstances.

In his article in the Stanford Encyclopedia of Philosophy Paul Bartha suggests a
Candidate Analogical Inference Rule based on the slogan 'probabilistic support
for future similarities is an increasing function of known past similarities'

A natural attempt here is to take the 'similarity' to apply between constants:

The General Analogy Principle, GAP

For $\vec{a} = \langle a_3, a_4, \ldots, a_k \rangle$ and $\psi(a_1, \vec{a}), \phi(a_1, \vec{a}) \in SL$,

$$w(\phi(a_2, \vec{a}) \,|\, \psi(a_1, \vec{a}) \wedge \psi(a_2, \vec{a}) \wedge \phi(a_1, \vec{a})) \geq w(\phi(a_2, \vec{a}) \,|\, \phi(a_1, \vec{a}))$$

GAP then attempts to capture *analogical support by similarity of properties*

Theorem 5. *Let w be a probability function on L satisfying Px + SN (+ Ex).
Then w satisfies GAP just if*

$$w = c_0^{L_q} = 2^{-q}(w_{\langle 1,0,0,\ldots,0 \rangle} + w_{\langle 0,1,0,\ldots,0 \rangle} + \ldots w_{\langle 0,\ldots,0,0,1 \rangle}),$$

equivalently,

$$w\big(\forall x \, \alpha_i(x)\big) = 2^{-q} \quad \text{for } i = 1, 2, \ldots, 2^q.$$

A proof of this result may be found in [13].

In summary then it appears that these previous two ways of measuring analog-
ical support miss the target. There is however another approach which is both
more general (since it applies also to polyadic rather than just unary languages)
and more closely captures our intuitions about what constitutes an analogy:

The Counterpart Principle, CP

*For any $\theta \in SL_q$, if $\theta' \in SL_q$ is obtained by replacing some of the predicate and
constant symbols appearing in θ by (distinct) new ones not occurring in θ and
$\psi \in SL_q$ only mentions constant and predicate symbols common to both θ and
θ' then*

$$w(\theta \,|\, \theta' \wedge \psi) \geq w(\theta \,|\, \psi). \tag{4}$$

---

[9]Where we require w to be fixed when we replace a relation symbol throughout a sentence
by its negation.

The Counterpart Principle focuses on an alternative interpretation of analogy, *analogy by structural similarity*, the structural similarity of θ' to θ. We would claim that this is just the similarity that we appreciate in the birthday party example (and countless other examples of analogical reasoning, in particular those we are aware of in Navya-Nyaya Logic) and which in turn is the basis for the support it provides. The following result, see [12], [26] shows how widely CP is satisfied:

Theorem 6. *Let the probability function w on L_q satisfy ULi. Then w satisfies the Counterpart Principle, CP.*

Proof. Clearly we may assume that $w(\psi) > 0$, otherwise the required conclusion is trivial. Since w satisfies ULi let the corresponding family of probability functions consist of w^r on L_r for $r \in \mathbb{N}^+$. Then

$$w_\infty = \bigcup_{r=1}^\infty w_r$$

is a probability function on the infinite (unary) language $L_\infty = \{R_1, R_2, R_3, \ldots\}$ extending w and satisfying Ex and Px. Let θ, θ', ψ be as in the statement of CP. Since we are assuming Ex and Px we may assume that all the constant and predicate symbols appearing in θ which are common θ' are amongst $a_1, a_2, \ldots, a_k$, $R_1, R_2, \ldots, R_g$, and that the replacements are $a_{n+i} \mapsto a_{n+i+k}$ for $i = 1, \ldots, k$ and $R_{g+j} \mapsto R_{g+j+t}$ for $j = 1, \ldots, t$. So suppressing these common constant and predicate symbols we can write

$$\theta = \theta(a_{n+1}, a_{n+2}, \ldots, a_{n+k}, R_{g+1}, R_{g+2}, \ldots, R_{g+t}),$$

$$\theta' = \theta(a_{n+k+1}, a_{n+k+2}, \ldots, a_{n+2k}, R_{g+t+1}, R_{g+t+2}, \ldots, R_{g+2t}).$$

With this in place let

$$\theta_{i+1} = \theta(a_{n+ik+1}, a_{n+ik+2}, \ldots, a_{n+(i+1)k}, R_{g+it+1}, R_{g+it+2}, \ldots, R_{g+(i+1)t}) \in SL_\infty$$

so $\theta_1 = \theta$, $\theta_2 = \theta'$. Now define $\tau : QFSL_1 \to SL_\infty$ by

$$\tau(R_1(a_i)) = \theta_i, \quad \tau(\neg\phi) = \neg\tau(\phi), \quad \tau(\phi \wedge \eta) = \tau(\phi) \wedge \tau(\eta), \quad \text{etc.}$$

for $\phi, \eta \in QFSL_1$.

Let $v : QFSL_1 \to [0, 1]$ be defined by

$$v(\phi) = w_\infty(\tau(\phi) \,|\, \psi).$$

Then since w_∞ satisfies (P1-2) (on SL_∞) so does v (on $QFSL_1$). Also since w_∞ satisfies Ex + Px, for $\phi \in QFSL_1$, permuting the θ_i in $w(\tau(\phi) \,|\, \psi)$ will leave this value unchanged so permuting the a_i in ϕ will leave $v(\phi)$ unchanged. i.e. v satisfies Ex.

By Gaifman's Theorem v has an extension to a probability function on L_1 and it is easy to check that it still satisfies Ex. Hence it also satisfies PIR by Theorem 4. In particular then

$$v(R_1(a_1) \,|\, R_1(a_2)) \geq v(R_1(a_1)).$$

But since $\tau(R_1(a_1)) = \theta$, $\tau(R_1(a_2)) = \theta'$ this amounts to

$$w_\infty(\theta \,|\, \theta' \wedge \psi) \geq w_\infty(\theta \,|\, \psi)$$

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

$\square$

This theorem then provides a *rational* justification for analogical support by structural similarity. Given this one might then expect that the more constants and predicates θ and θ' have in common the greater this support. That indeed turns out to be the case, the following result is sketched in [26, p173]:

Theorem 7. *Let the probability function w on L_q satisfy ULi and let*

$$\theta = \theta(\vec{a_1}, \vec{a_2}, \vec{a_3}, \vec{R_1}, \vec{R_2}, \vec{R_3})$$
$$\theta' = \theta(\vec{a_1}, \vec{a_2}, \vec{a_4}, \vec{R_1}, \vec{R_2}, \vec{R_4})$$
$$\theta'' = \theta(\vec{a_1}, \vec{a_5}, \vec{a_6}, \vec{R_1}, \vec{R_5}, \vec{R_6})$$

and $\psi = \psi(\vec{a_1}, \vec{R_1})$ where the $\vec{a_i}, \vec{R_j}$ are all disjoint. Then

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

Interestingly the proof of this perhaps rather unsurprising theorem seems to require results from *Polyadic Inductive Logic*, that is where we allow not just unary predicate symbols in our language but also relation symbols of all arities. What is more Theorems 6 and 7 continue to hold in polyadic inductive logic provided we enhance ULi to the obvious polyadic version Li.

It seems interesting to note that, as exemplified by the two versions of 'analogy' treated here, that to date the analogy principles which have been investigated are either consequences of existing well accepted symmetry principles (as was the case with the counterpart Principle and Ex) or else only hold for some very few probability functions which hardly justify being termed rational, see for example [11], [13]. At this stage then it seems, in view of our earlier remarks, that the notions of relevance and analogy (if indeed these are not already essentially the same thing) are simply derivatives of symmetry and irrelevance and that these two are the basic 'building blocks' of rationality in this context.

Theorems 6 and 7 seem to be the tip of an iceberg, which we (Alena Vencovská) currently plan to investigate. For example, assuming ULi throughout, in which

cases do we have strict inequality in the Counterpart Principle? This clearly cannot hold if θ is a contradiction or a tautology and there are other examples too of 'constant' sentences θ which always give equality in the Counterpart Principle for any probability function satisfying Ex + Px. We have a complete characterization of such 'constant sentences' but what we do not yet have is a complete characterization of the probability functions which give a strict inequality in the Counterpart Principle for all but these constant sentences. [For more details see [12].]

A second issue is the combined effect of multiple analogies, or for example replacing θ' by some ϕ' where $\phi \models \theta$. Indeed it seems possible that there is a 'logic of analogous reasoning' to be unearthed here.

References

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

[2] Carnap, R. & Jeffrey, R.C. eds., *Studies in inductive logic and probability,* Volume I, University of California Press, 1971.

[3] Carnap, R., A basic system of inductive logic, in *Studies in Inductive Logic and Probability*, Volume II, ed. R. C. Jeffrey, University of California Press, 1980, pp7-155.

[4] 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.

[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., *Reasoning by Analogy in Inductive Logic*, Doctorial Thesis, Manchester, UK, July 2013. Available at http://www.maths.manchester.ac.uk/ jeff/

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

[12] Hill, A. & Paris, J.B., The Counterpart Principle of Analogical Support by Structural Similarity, to appear in *Erkenntnis.*

[13] Howarth, E., Paris, J.B. & Vencovská, A., An examination of the SEP Candidate Analogical Inference Rule within Pure Inductive Logic, to be submitted to the *Journal of Applied Logic.*

[14] Humburg, J.: The principle of instantial relevance, in *Studies in Inductive Logic and Probability (Volume I),* Eds. R.Carnap & R.C.Jeffrey, University of California Press, Berkeley and Los Angeles, 1971.

[15] Hume,D. *An Enquiry Concerning Human Understanding,* 1748.

[16] Johnson, W.E., Probability: The deductive and inductive problems, *Mind*, **41**:409-423, 1932.

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

[18] Kließ, M.S. & Paris, J.B., Predicate Exchangeability and Language Invariance in Pure Inductive Logic, submitted to *Logique et Analyse.*

[19] Maher, P., Probabilities for multiple properties: The models of Hesse, Carnap and Kemeny, *Erkenntnis*, 2001, **55**:183-216.

[20] Maher, P., A Conception of Inductive Logic, *Philosophy of Science*, 2006, **73**:513-520.

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

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

[23] 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.

[24] Paris, J.B., *Guangzhou Winter School Notes on Pure Inductive Logic 2010.* See `http://www.maths.manchester.ac.uk/~jeff/`

[25] Paris, J.B. & Vencovská, A., A Note on the Inevitability of Maximum Entropy, *International Journal of Approximate Reasoning*, (1990) pp.183-224, Vol.4, No.3.

[26] 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, 2014.

[27] Romeijn, J-W., Analogical Predictions for Explicit Simmilarity, *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.

[29] Waterhouse, P.J., *Probabilistic Relationships, Relevance and Irrelevance within the Field of Uncertain Reasoning*, Doctorial Thesis, The University of Manchester, Manchester, UK, 2007. Available at http://www.maths.manchester.ac.uk/ jeff/theses/pwthesis.pdf