17.12.10

Pure Inductive Logic

Winter School in Logic, Guangzhou, China, 2010

Jeff Paris,
School of Mathematics,
University of Manchester,
Manchester M13 9PL, UK.

jeff.paris@manchester.ac.uk

Introduction

Before a football match can begin the tradition is that the referee tosses a coin and one of the captains calls, heads or tails, whilst the coin is in the air. If the captain gets it right s/he picks which end to start playing at, or alternatively to have the kick off. There never seems to be an issue of which captain actually makes this call (otherwise we would have to toss a coin and make a call to decide who makes the call, and in turn toss a coin and make a call to decide who makes that call and so on) since it seems clear that this is fair, that both captains give equal probability to the coin landing heads as to it landing tails no matter which of them calls it. The obvious explanation for this is that both captains are appealing to the *symmetry* of the situation.

At the same time they are, it seems, making the assumption that all the other information they possess about the situation, for example the weather, the gender of the referee, even past successes at coin calling, is *irrelevant*, at least if it doesn't involve some specific knowledge about this particular coin or tosser (i.e. the referee). Of course if we knew that on the last 8 occasions on which this particular referee had tossed up the result had been heads we might well consider that *was relevant.*

Forming beliefs, or subjective probabilities, in this way by considering symmetry, irrelevance, relevance, can be thought of as *logical* or *rational* inference. This is something different from statistical inference. The perceived fairness of the coin toss is clearly not based on the captains' knowledge of a long run of past tosses by the referee which have favored heads close to half the time. Indeed it is conceivable that this long run frequency might not give an average of close to half heads, maybe this dexterous referee has developed a skill for influencing the spin and trajectory. Nevertheless even if the captains knew that the referee was so gifted unless they were privy to which side of the coin the referee favored they would surely still consider the process as fair.

This illustrates another feature of probabilities inferred on logical grounds: they certainly need not agree with the 'true', or long term frequency probability, if this even exists, and of course in many situations in which we form subjective probabilities no such 'true' probability does exist. For example when assigning odds in a horse race.

The aim of this short course is to provide an introduction to this logic of assigning probabilities on the basis of notions such as symmetry, irrelevance, relevance on which the assignment appears to depend. Much has already been written by philosophers on these matters and doubtless much still remains to be said. However our approach here will be that of mathematical, rather than philosophical, logicians. So instead of spending a significant time discussing these notions at length in the context of specific examples we shall largely consider ways in

which they might be given a purely mathematical formulation and then devote our main effort to describing some of the mathematical and logical consequences which ensue.

This difference in approach between the mathematical and philosophical is reflected in the name *Pure Inductive Logic*, used in a similar fashion to the division between Pure and Applied Mathematics. Starting from the founders of Inductive Logic, the philosophers W.E.Johnson around 1930 and independently R.Carnap in the early 1940's, see [17],[3],[4],[5],[6], philosophers had seen this subject as potentially applicable, providing rules for uncertain reasoning such the propositional and predicate calculi provide for categorical reasoning. It was this pragmatic aspect which was to be subsequently undermined by N.Goodman's *Grue Paradox*, see the Appendix or [14],[15], leading to a widespread abandonment of Carnap's programme within philosophy. However this 'paradox' is no barrier at all to the development of inductive logic as *pure mathematical logic*.

In this way then we are promoting, or at least recognizing, a new area of mathematical logic, namely the title of this course. It is not philosophy as such but there are close connections. Firstly most of the logical, aka rational, principles we consider are motivated by philosophical considerations, frequently having an already established presence and literature within that subject. Secondly we would hope that the mathematical results included here may feed back and contribute to the continuing debates within philosophy, if only by saying that *if* you subscribe to A, B, C *then* you must, by dint of mathematical proof, accept D.

There is a parallel here with Set Theory. In that case we propose axioms based on our intuitions concerning the nature of sets and then investigate their consequences. These axioms have philosophical content and considering this is part of the picture but so also is drawing out their mathematical relationships and consequences. And as we go deeper into the subject we are led to propose or investigate axioms which initially might not have entered our minds, not least because we may well not have possessed the language or notions to even express them.

And at the end of the day most of us would like to think that discoveries in set theory were telling us something about the universe of sets, or at least about possible universes of sets, and thus feeding back into the philosophical debate (and not just generating mathematics 'because it is there'!).

Context

For the mathematical setting we need to make the formalism completely clear. Whilst there are various possible choices here the language which seems best for our study, and corresponds to most of the literature, including Carnap's, is where

we work with a first order language L with *unary*[1] relation symbols $R_1, R_2, \ldots, R_q$ and constants a_n for $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$, and no function symbols nor (in general) equality. The intention here is that the a_i name all the individuals in some population though there is no prior assumption that they necessarily name different individuals. We identify L with the set $\{R_1, R_2, \ldots, R_q\}$. Let SL denote the set of first order sentences of this language L and $QFSL$ the quantifier free sentences of this language. Similarly we use $FL, QFFL$ for the formulae of L. *We shall throughout adopt the convention that if we write a formula θ of L as $\theta(\vec{x})$ etc. then there are no occurrences of the constant symbols a_i in θ unless otherwise stated.*

Let $\mathcal{T}$ denote the set of structures for L with universe $\{a_1, a_2 \ldots\}$, with the obvious interpretation of the a_i as a_i itself.

To capture the underlying problem that Inductive Logic aims to address we can now imagine an agent who inhabits some structure $M \in \mathcal{T}$ but knows nothing about what is true in M. Then the problem is,

> $\mathcal{Q}$: *In the situation of zero knowledge, logically, or rationally, what belief should our agent give to a sentence $\theta \in SL$ being true in M?*

There are several terms in this question which need explaining. Firstly 'zero knowledge' means that the agent has no intended interpretation of the a_i nor the R_j. To mathematicians this seems a perfectly easy idea to accept, we already do it effortlessly when proving results about, say, an arbitrary group. In these cases all you can assume is the axioms and you are not permitted to bring in new facts because they happen to hold in some particular group you have in mind. Unfortunately outside of mathematics this sometimes seems to be a particular difficult idea to embrace and much confusion has found its way into the folklore as a result.[2]

In a way this is at the heart of the difference between the 'Pure Inductive Logic' proposed here as mathematics and the 'Applied Inductive Logic' of Carnap as philosophy. For many philosophers would argue that in this latter the language is intended to carry with it an interpretation, that without it one is doing pure mathematics not philosophy. It is the reason why Grue is a paradox in philosophy and simply an invalid argument in mathematics. However we all need to be on our guard when it comes to allowing interpretations to slip in subconsciously.

Second unexplained terms are 'logical' and its synonym (as far as this text is concerned) 'rational'. In this case we shall offer no definition, they are to be

[1] Only over the past decade or so has Inductive Logic embraced binary, ternary etc. relations. In this introductory course we shall restrict ourselves, as the subject founders Johnson and Carnap did, to purely unary languages.

[2] See for example [24] and the issue of the representation dependence of maxent.

taken as intuitive, something we recognize when we see it without actually being able to give it a definition.[3] This will not be a great problem for our purpose is to propose and mathematically investigate principles for which it is enough that we may simply *entertain* the idea that they are logical or rational. The situation parallels that of the intuitive notion of an 'effective process' in recursion theory, and similarly we may hope that our investigations will ultimately lead to a clearer understanding.

The third unexplained term above is 'belief'. For the present we shall identify belief, or more precisely degree of belief, with probability and only later provide a justification, the Dutch Book Argument, for this identification. The main reason for proceeding in this way is that in order to give this argument in full we actually need to have already developed some of the apparatus of probability functions, a task we now move on to.

Probability Functions

Definition

A function $w : SL \to [0, 1]$ is a probability function on SL if for all $\theta, \phi, \exists x\, \psi(x) \in SL$,[4]

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

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

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

Condition P3 is often referred to as *Gaifman's Condition*, see [12], and is a special addition to the conventional conditions P1, P2 appropriate to this context. It intends to capture the idea that the $a_1, a_2, \ldots$ exhaust the universe.

A particularly simple example of a probability function is the function $V_M : SL \to \{0, 1\}$, where $M \in \mathcal{T}$, defined by

$$V_M(\theta) = \begin{cases} 1 & \text{if } M \vDash \theta, \\ 0 & \text{otherwise,} \end{cases}$$

In turn, since P1-3 are all linear, convex sums of probability functions are also probability functions. So for $M_i \in \mathcal{T}$ and $a_i \geq 0$ such that $\sum_i a_i = 1$, $\sum_i a_i V_{M_i}$ is a probability function.

More generally for $\theta \in SL$ let

$$[\theta] = \{\, M \in \mathcal{T} \mid M \vDash \theta \,\}$$

[3] Just as we can recognize the smell of an onion despite having no words to describe it.

[4] Here $\psi(x)$ may mention constants a_i.

let $\mathcal{B}$ be the σ-algebra of subsets of $\mathcal{T}$ generated by these subsets $[\theta]$ and let μ be a countable additive measure[5] on $\mathcal{B}$. Then it is straightforward to check that

$$w = \int V_M \, d\mu(M)$$

defines a probability function on SL.

In fact as we shall see later the converse to this result holds. That is, for any probability function w on L there is a countably additive measure μ_w on $\mathcal{B}$ such that for any $\theta \in SL$,

$$w(\theta) = \int_{\mathcal{T}} V_M(\theta) \, d\mu_w(M). \tag{1}$$

Simple properties of probability functions

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

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

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

(c) $\theta \vDash \phi \;\Rightarrow\; w(\theta) \leq w(\phi)$.

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

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

Proof (a) We have that $\vDash \theta \vee \neg\theta$ and $\theta \vDash \neg\neg\theta$ so by P1 and P2,

$$1 = w(\theta \vee \neg\theta) = w(\theta) + w(\neg\theta).$$

(b) If $\vDash \neg\theta$ then $w(\neg\theta) = 1$ by P1 so from (a), $w(\theta) = 0$.

(c) If $\theta \vDash \phi$ then $\neg\phi \vDash \neg\theta$ so from P2, (a) and the fact that w takes values in $[0, 1]$,

$$1 \geq w(\neg\phi \vee \theta) = w(\neg\phi) + w(\theta) = 1 - w(\phi) + w(\theta)$$

from which the required inequality follows.

(d) If $\theta \equiv \phi$ then $\theta \vDash \phi$ and $\phi \vDash \theta$. By (c), $w(\theta) \leq w(\phi)$ and $w(\phi) \leq w(\theta)$ so $w(\theta) = w(\phi)$.

(e) Since $\theta \vee \phi \equiv \theta \vee (\neg\theta \wedge \phi)$ and $\theta \vDash \neg(\neg\theta \wedge \phi)$ P2 and (d) give

$$w(\theta \vee \phi) = w(\theta \vee (\neg\theta \wedge \phi)) = w(\theta) + w(\neg\theta \wedge \phi). \tag{2}$$

[5]All measures will be assumed to be normalized unless otherwise indicated.

Also $\phi \equiv (\theta \wedge \phi) \vee (\neg\theta \wedge \phi)$ and $\theta \wedge \phi \models \neg(\neg\theta \wedge \phi)$ so by P2 and (d),

$$w(\phi) = w((\theta \wedge \phi) \vee (\neg\theta \wedge \phi)) = w(\theta \wedge \phi) + w(\neg\theta \wedge \phi). \tag{3}$$

Eliminating $w(\neg\theta \wedge \phi)$ from (2), (3) now gives

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

■

For future use it is worthwhile pointing out here that nowhere in the proof of Proposition 1 did we use the property P3, each of (a)-(e) holds even if we only assume P1 and P2. Indeed if we restrict θ, ϕ here to quantifier free sentences then we only need assume P1 and P2 for quantifier free sentences.

Proposition 2 *For $\exists x\, \psi(x) \in SL$ and $w : SL \to [0,1]$ satisfying* P1, P2 *condition* P3 *is equivalent to*

$$\text{P3}' \quad w(\exists x\, \psi(x)) = \sum_{i=1}^{\infty} w\left(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)\right).$$

Proof Since

$$\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n) \;\equiv\; \bigvee_{j=1}^{n} \left(\psi(a_j) \wedge \neg \bigvee_{i=1}^{j-1} \psi(a_i)\right)$$

by Proposition 1(d) and the remark following the proof of the proposition,

$$\begin{aligned} w(\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n)) &= w\left(\bigvee_{j=1}^{n} \left(\psi(a_j) \wedge \neg \bigvee_{i=1}^{j-1} \psi(a_i)\right)\right) \\ &= \sum_{j=1}^{n} w\left(\psi(a_j) \wedge \neg \bigvee_{i=1}^{j-1} \psi(a_i)\right) \end{aligned}$$

by repeated use of P2 since the disjuncts here are all disjoint. The required equivalence of P3, P3′ follows. ■

Condition P3 is expressed in terms of the probability of existential sentences. However it could equally well have been expressed in terms of universal sentences as the next proposition shows.

Proposition 3 *Let $w : SL \to [0,1]$ satisfy* P1, P2. *Then condition* P3 *is equivalent to:*

$$w(\forall x\, \eta(x)) = \lim_{n\to\infty} w\left(\bigwedge_{i=1}^{n} \eta(a_i)\right) \tag{4}$$

for $\forall x\, \eta(x) \in SL$.

Proof Assume P3. Then

$$\begin{aligned} w(\forall x\, \eta(x)) &= 1 - w(\exists x\, \neg\eta(x)) \quad \text{by Proposition 1(a)(d),} \\ &= 1 - \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \neg\eta(a_i)\right) \quad \text{by P3} \\ &= \lim_{n\to\infty} \left(1 - w\left(\bigvee_{i=1}^{n} \neg\eta(a_i)\right)\right) \\ &= \lim_{n\to\infty} w\left(\bigwedge_{i=1}^{n} \eta(a_i)\right) \quad \text{by Proposition 1(a)(d).} \end{aligned}$$

The converse follows by simply reversing this argument. ■

Exercise Give a proof of Suppes Lemma:

Lemma 4 *Let $\theta_1, \theta_2, \ldots, \theta_n \in SL$ and w a probability function on L. Then*

$$w\left(\bigwedge_{i=1}^{n} \theta_i\right) \geq \sum_{i=1}^{n} w(\theta_i) - (n-1).$$

In particular if $w(\theta_i) = 1$ for $i = 2, 3, \ldots, n$, $(n \geq 1)$ then

$$w\left(\bigwedge_{i=1}^{n} \theta_i\right) = w(\theta_1).$$

The Dutch Book Argument

Having derived some of the basic properties of probability functions we will now take a short diversion to give what we consider to be the main argument, namely the Dutch Book argument, in favor of an agent's 'degrees of belief' satisfying P1-3, and hence being identified with a probability function, albeit a *subjective* probability since it is ostensibly the property of the agent in question. Of course this could really be said to be an aside to the purely mathematical study of Inductive Logic and hence could be dispensed with. The advantage of considering this argument however is that by linking belief and subjective probability it better enables us to appreciate and translate into mathematical formalism the many rational principles we shall later encounter.

The idea of the Dutch Book argument is that it identifies 'belief' with willingness to bet. So suppose, as in the context of Inductive Logic explained above we have an agent inhabiting some possibly unknown structure $M \in \mathcal{T}$ (which one imagines

will eventually be revealed to decide the wager) and that $\theta \in SL$, $0 \leq p \leq 1$ and for a stake $s > 0$ the agent is offered a choice of one of two wagers:

(Bet1$_p$) Win $s(1-p)$ if $M \models \theta$, lose sp if $M \not\models \theta$.

(Bet2$_p$) Win sp if $M \not\models \theta$, lose $s(1-p)$ if $M \models \theta$.

If the agent would not be happy to accept Bet1$_p$ we assume that it is because the agent thinks that the bet is to his/her disadvantage and hence to the advantage of the bookmaker.[6] But in that case Bet2$_p$ allows the agent to swap roles with the bookmaker so s/he should now see that bet as being to his/her advantage, and hence acceptable. In summary then, we may suppose that for any $0 \leq p \leq 1$ at least one of Bet1$_p$ and Bet2$_p$ is acceptable to the agent.

Now suppose that Bet1$_p$ was acceptable to the agent and $0 \leq q < p$. Then Bet1$_q$ should also be acceptable to the agent[7] since it would result for a greater win, $s(1-q)$, than Bet1$_p$ if θ turns out to be true in M and a smaller loss, sq, than Bet1$_p$ if θ turns out to be false. Similarly if Bet2$_p$ is acceptable to the agent and $p < q \leq 1$ then Bet2$_q$ should be acceptable to the agent.

From this it follows that those $p \in [0,1]$ for which Bet1$_p$ is acceptable to the agent form an initial segment of the interval $[0,1]$, those for which Bet2$_p$ is acceptable form a final segment and every p is in one or other of these segments, possibly even both.

Define $Bel(\theta)$ to be the supremum of those $p \in [0,1]$ for which Bet1$_p$ is acceptable to the agent. We can argue that $Bel(\theta)$ is a measure of the agent's willingness to bet on θ and in turn take this to quantify the agent's belief that θ is true in M. For if the agent strongly believes that θ is true then s/he would be willing to risk a small potential gain of $s(1-p)$ against a large potential loss of sp, simply because s/he strongly expects that gain, albeit small. In other words the agent would favor Bet1$_p$ even for p quite close to 1. From this viewpoint then $Bel(\theta)$ represents the top limit of the agent's belief in θ.

We now impose a further rationality requirement on the agent: That s/he does not allow a (possibly infinite) set of simultaneous bets each of which is acceptable to him/her but whose combined effect is to cause the agent certain loss no matter what the ambient structure $M \in \mathcal{T}$ turns out to be. In common parlance that the agent cannot be 'Dutch Booked'.[8]

To formalize this idea first observe that if the agent accepts Bet1$_p$ s/he will in the event of the ambient structure being M gain in pounds

$$s(1-p)V_M(\theta) - sp(1 - V_M(\theta)) = s(V_M(\theta) - p)$$

[6]For this purpose we may suppose that we have selected an agent who does not suffer from any morally aversion to gambling or the like.

[7]Assuming that s/he is 'rational'.

[8]The word 'Dutch' here, though originally derived from the nation, simply means 'queer', as in 'double dutch'.

where V_M was defined on page 5 and loss is identified with negative gain. Clearly in Bet2$_p$ the gain is minus this, so $-s(V_M(\theta) - p)$.

So the agent could certainly be Dutch Booked if there were countable sets A, B and sentences θ_i, stakes $s_i > 0$, $p_i \in [0, Bel(\theta_i))$ for $i \in A$, and sentences ϕ_j, stakes $t_j > 0$, $q_j \in (Bel(\phi_j), 1]$ for $j \in B$, and $K > 0$ such that

$$\sum_{i \in A} s_i(V_M(\theta_i) - p_i) + \sum_{j \in B}(-t_j)(V_M(\phi_j) - q_j) < 0 \tag{5}$$

and

$$\sum_{i \in A} s_1 V_M(\theta_i), \sum_{i \in A} s_i p_i, \sum_{j \in B} t_j V_M(\phi_j), \sum_{j \in B} t_j q_j < K \tag{6}$$

for all $M \in \mathcal{T}$. For in this case the agent would accept Bet1$_{p_i}$ for θ_i at stake s_i for each $i \in A$ and similarly would accept Bet2$_{q_j}$ for ϕ_j at stake t_j for each $j \in B$, and by condition (6) it would be feasible for the agent to so do, but the result of simultaneously playing all these bets would, by (5), be negative no matter what $M \in \mathcal{T}$ was.

We now show that imposing the condition that no such p_i, s_i, θ_i etc. exist forces Bel to satisfy P1-3 and hence to be a probability function according to our definition. The original proof of this is due to de Finetti, [10], for P1-2. Williamson, [29, Theorem 5.1], gives a Dutch Book argument for P3 though the one we present here is somewhat different.

Theorem 5 *Suppose that for $Bel : SL \to [0, 1]$ there are no countable sets A, B, sentences $\theta_i \in SL$, $p_i \in [0, Bel(\theta_i))$, stakes s_i for $i \in A$ etc. such that (5) holds. Then Bel satisfies P1-3.*

Proof For (P1) suppose that $\theta \in SL$ and $\vDash \theta$ but $Bel(\theta) < 1$. Then for $Bel(\theta) < q < 1$ the agent accepts Bet2$_q$. But since $V_M(\theta) = 1$ for all $M \in \mathcal{T}$ we have that with stake 1,

$$(-1)(V_M(\theta) - q) = q - 1 < 0$$

which gives an instance of (5), contradiction.

Now suppose that P2 fails say $\theta, \phi \in SL$ are such that $\theta \vDash \neg\phi$ but

$$Bel(\theta) + Bel(\phi) < Bel(\theta \vee \phi).$$

Then $\theta \vDash \neg\phi$ forces that at most one of θ, ϕ can be true in any $M \in \mathcal{T}$ so

$$V_M(\theta \vee \phi) = V_M(\theta) + V_M(\phi).$$

Pick $p > Bel(\theta)$, $q > Bel(\phi)$, $r < Bel(\theta \vee \phi)$ such that $p + q < r$. Then with stakes 1,1,1,

$$(-1)(V_M(\theta) - p) + (-1)(V_M(\phi) - q) + (V_M(\theta \vee \phi) - r) = (p + q) - r < 0$$

giving an instance of (5) and contradicting our assumption. A similar argument when

$$Bel(\theta) + Bel(\phi) > Bel(\theta \vee \phi)$$

shows that this cannot hold either so we must have equality here.

Finally suppose that $\exists x\, \psi(x) \in SL$. By Proposition 2 and the fact that we have already proved that P1, P2 hold for Bel, it is enough to derive a contradiction from the assumption that

$$\sum_{i=1}^{\infty} Bel(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)) \neq Bel(\exists x\, \psi(x)).$$

Notice that since the sentences on the left hand side here are disjoint both sides here are bounded by 1. Suppose first that we had $<$ here. Then we could pick

$$p_n > Bel(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)) \quad \text{for } n = 1, 2, 3, \ldots$$

and $r < Bel(\exists x\, \psi(x))$ with $\sum_{n=1}^{\infty} p_n < r$. Since for $M \in \mathcal{T}$,

$$V_M(\exists x\, \psi(x)) = \sum_{n=1}^{\infty} V_M(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i))$$

we get, as with the argument above for P2, that for all stakes 1,

$$(V_M(\exists x\, \psi(x)) - r) + \sum_{n=1}^{\infty} (-1)(V_M(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)) - p_n) = -r + \sum_{n=1}^{\infty} p_n < 0,$$

giving an instance of (5) (and (6) in contradiction to our assumption. The case where we have $>$ in place of $<$ here is proved similarly. ■

Theorem 5 tells us then that the beliefs of a rational agent, rational in the sense of not allowing him/herself to be Dutch Booked, may be quantified as a probability function. That however raises the question whether there might be additional properties P4,P5 etc. of Bel which follow from there being no instances of (5). The answer to this is 'No', as the next theorem shows.

Theorem 6 *Suppose that $w : SL \to [0, 1]$ is a probability function. Then betting according to w the agent cannot be Dutch Booked. (Where in the case $p = w(\theta)$ the agent is allowed to choose either $Bet1_p$ or $Bet2_p$.)*

Proof Suppose on the contrary that there were countable sets A, B and some sentences θ_i for $i \in A$, stakes $s_i > 0$ and $p_i \in [0, w(\theta_i)]$ and sentences ϕ_j for $j \in B$ and stakes $t_j > 0$ and $q_j \in [w(\phi_j), 1]$, and $K > 0$ such that (5), (6) held

(now with closed intervals rather than half open). Then using the measure μ_w introduced earlier in (1), from (6)

$$\begin{aligned}
&\sum_{i\in A} s_i(w(\theta_i) - p_i) + \sum_{j\in B}(-t_j)(w(\phi_j) - q_j) \\
&\quad = \sum_{i\in A} s_i(\int V_M(\theta_i)\, d\mu_w(M) \ - p_i) + \sum_{j\in B}(-t_j)(\int V_M(\phi_j)\, d\mu_w(M) \ - q_j) \\
&\quad = \int \sum_{i\in A} s_i(V_M(\theta_i) - p_i) + \sum_{j\in B}(-t_j)(V_M(\phi_j) - q_j)\, d\mu_w(M),
\end{aligned}$$

by the Bounded Convergence Theorem, see for example [2]. By (5) the last, and hence also the first, expressions here are negative. But that is impossible since $w(\theta_i) - p_i \geq 0$ and $w(\phi_j) - q_j \leq 0$. ■

Specifying Probability Functions

On the face of it it might appear that because of the great diversity of sentences in SL 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, [12]:

Theorem 7 *Suppose that $w : QFSL \to [0,1]$ satisfies (P1) and (P2) for $\theta, \phi \in QFSL$. Then w has a unique extension to a probability function on SL satisfying (P1),(P2),(P3) for any $\theta, \phi, \exists x\, \psi(x) \in SL$.*

Proof Let w be as in the statement of the theorem. For $\theta \in QFSL$ the subsets

$$[\theta] = \{\, M \in \mathcal{T} \mid M \vDash \theta \,\}$$

of $\mathcal{T}$ form an algebra $\mathcal{A}$ of sets and μ_w defined by

$$\mu_w([\theta]) = w(\theta) \qquad \text{for } \theta \in QFSL$$

is easily seen to be a finitely additive measure on this algebra.

Indeed μ_w is (trivially) σ-finitely additive. For suppose $\theta, \phi_i \in QFSL$ for $i \in \mathbb{N}$ and

$$\bigcup_{i\in\mathbb{N}} [\phi_i] = [\theta]. \tag{7}$$

Then it must be the case that for some finite n

$$\bigcup_{i\leq n} [\phi_i] = [\theta],$$

otherwise

$$\{\, \neg\phi_i \,|\, i \in \mathbb{N} \,\} \cup \{\theta\}$$

would be finitely satisfiable and hence by Compactness would be satisfiable in some $M \in \mathcal{T}$, contradicting (7).

Hence by Carathéodory's Extension Theorem (see for example [1]) there is a unique extension μ_w^+ of μ_w defined on the σ-algebra $\mathcal{B}$ generated by $\mathcal{A}$. Notice that since

$$\begin{aligned}
[\exists x\, \psi(x)] &= \{\, M \in \mathcal{T} \,|\, M \vDash \exists x\, \psi(x) \,\} \\
&= \{\, M \in \mathcal{T} \,|\, M \vDash \psi(a_i), \text{ some } i \in \mathbb{N}^+ \,\} \\
&= \bigcup_{i \in \mathbb{N}^+} \{\, M \in \mathcal{T} \,|\, M \vDash \psi(a_i) \,\} \\
&= \bigcup_{i \in \mathbb{N}^+} [\psi(a_i)] \qquad (8)
\end{aligned}$$

and $\mathcal{B}$ is closed under complements and countable unions the algebra $\mathcal{B}$ contains all the sets $[\theta]$ for $\theta \in SL$.

Now define w^+ on SL by

$$w^+(\theta) = \mu_w^+([\theta]).$$

Then since μ_w^+ extends μ_w, w^+ extends w. Since μ_w^+ is a measure w^+ satisfies P1-2 and also P3 from (8) and the fact that μ_w^+ is countably additive.

Finally w^+ must be the unique extension of w on SL satisfying P1-3. For if there was another such probability function, u say, then the measure μ_u on the algebra of sets $[\theta]$ for $\theta \in QFSL$ would have a countably additive extension μ_u^+ to the σ algebra generated by these $[\theta]$. But clearly that algebra would have to be $\mathcal{B}$ so μ_u^+ would have to agree with μ_w^+ (because of its uniqueness) and in turn u would have to agree with w on SL.

■

We are now in a position to justify the assertion (1) which we made earlier.

Corollary 8 *Suppose that w is a probability function on L. Then there is a countably additive measure μ_w on the algebra $\mathcal{B}$ of subsets of $\mathcal{T}$ such that for $\theta \in SL$,*

$$w(\theta) = \int_{\mathcal{T}} V_M(\theta)\, d\mu_w(M).$$

Proof If we start with $u = w\restriction QFSL$ then the proof of Theorem 7 shows that there is a countably additive measure μ_u^+ on $\mathcal{B}$ such that for $\theta \in SL$,

$$u^+(\theta) = \mu_u^+([\theta]) = \int_{\mathcal{T}} W_M(\theta)\, d\mu_u^+(M).$$

But by uniqueness u^+ must be the same thing as w. ■

From Gaifman's Theorem 7 it follows that to specify a probability function on SL it is enough to say how it acts on the quantifier free sentences. We now explain how even that task can be simplified.

As usual let L be our default language with unary relation symbols $R_1, R_2, \ldots, R_q$. An *atom* $\alpha(x)$ of L is a formula of the form

$$\pm R_1(x) \wedge \pm R_2(x) \wedge \pm R_3(x) \wedge \ldots \wedge \pm R_q(x)$$

where $\pm R$ means one of R, $\neg R$.

Let $\alpha_1(x), \alpha_2(x), \ldots, \alpha_{2^q}(x)$ list the atoms of L. Clearly these atoms are disjoint and exhaustive and for a_i a constant of L everything there is to know about a_i is determined by the unique atom which a_i satisfies.

Similarly for distinct constants $b_1, b_2, \ldots, b_m$ coming from $a_1, a_2, \ldots$ everything there is to know about the b_i is determined by the unique *State Description*

$$\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$$

that the b_i satisfy.

We allow here the possibility that $m = 0$ in which case the sole state description for these constants is a tautology, for which we use the symbol $\top$.

As an example here, if L has just the unary relation symbols R_1, R_2 then the atoms of L are

$$R_1(x) \wedge R_2(x), \quad R_1(x) \wedge \neg R_2(x), \quad \neg R_1(x) \wedge R_2(x), \quad R_1(x) \wedge \neg R_2(x)$$

and a state description for a_1, a_3 could be

$$(R_1(a_1) \wedge \neg R_2(a_1)) \wedge (R_1(a_3) \wedge R_2(a_3)).$$

We shall use upper case Θ, Φ, Ψ etc. for state descriptions and, since we will only be interested in state descriptions up to logical equivalence, identify two state descriptions if they agree up to the order of their conjuncts.

By the Disjunctive Normal Form Theorem any $\theta(b_1, b_2, \ldots, b_m) \in QFSL$ is logically equivalent to a disjunction of state descriptions for $b_1, b_2, \ldots, b_m$, say

$$\theta(\vec{b}) \equiv \bigvee_{\Theta(\vec{b}) \in S} \Theta(\vec{b})$$

for some set S of state descriptions. Hence, since state descriptions for $b_1, b_2, \ldots, b_m$ are exclusive, using (P2) repeatedly,

$$w(\theta(\vec{b})) = \sum_{\Theta(\vec{b}) \in S} w(\Theta(\vec{b})). \tag{9}$$

From this it follows that the probability function w is determined on $QFSL$, and hence on all of SL by Theorem 7, by its values on state descriptions. Indeed since in (9) not all the $b_1, b_2, \ldots, b_m$ need actually appear in $\theta(\vec{b})$ to determine w it is enough to know w just on the state descriptions for $a_1, a_2, \ldots, a_m$, $m \in \mathbb{N}$.

Conversely if the function w is defined on the state descriptions $\Theta(a_1, a_2, \ldots, a_m)$ to satisfy:

(i) $w(\Theta(a_1, a_2, \ldots, a_m)) \geq 0$,

(ii) $w(\top) = 1$,

(iii) $$w(\Theta(a_1, a_2, \ldots, a_m)) = \sum_{\Phi(a_1,\ldots,a_{m+1}) \models \Theta(a_1,\ldots,a_m)} w(\Phi(a_1, a_2, \ldots, a_{m+1})),$$

(and notice that these are all properties that do hold for a probability function on L) then w extends to a probability function on $QFSL$, and hence on SL, by setting (unambiguously by (iii))

$$w(\theta(b_1, b_2, \ldots, b_m)) = \sum_{\Theta(a_1,\ldots,a_k) \models \theta(b_1,\ldots,b_m))} w(\Theta(a_1, a_2, \ldots, a_k))$$

where k is sufficiently large that all of the b_i are amongst $a_1, a_2, \ldots, a_k$.

Conditional Probability

Given a probability function w on SL and $\phi \in SL$ with $w(\phi) > 0$ we define the *conditional probability function* $w(. \,|\, \phi) : SL \to [0, 1]$ (said w conditioned on ϕ) by

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

Proposition 9 *Let w be a probability function on SL, $\phi \in SL$ and $w(\phi) > 0$. Then $w(. \,|\, \phi)$ is a probability function and $w(\theta \,|\, \phi) = 1$ whenever $\phi \models \theta$.*

Proof To show P1 suppose that $\models \theta$. Then $\phi \equiv \theta \wedge \phi$ so $w(\theta \wedge \phi) = w(\phi)$ by Proposition 1(d) and in turn $w(\theta \,|\, \phi) = 1$. Notice that all we really need for this argument is the logical equivalence $\phi \equiv \theta \wedge \phi$ so since it is enough to have $\phi \models \theta$ for this to hold we immediately get the final part of the proposition too.

For P2 suppose that $\theta \models \neg\eta$. Then $\theta \wedge \phi \models \neg(\eta \wedge \phi)$ so since

$$(\theta \vee \eta) \wedge \phi) \equiv (\theta \wedge \phi) \vee (\eta \wedge \phi),$$

$$\begin{aligned} w((\theta \vee \eta) \wedge \phi) &= w((\theta \wedge \phi) \vee (\eta \wedge \phi)) \quad \text{by Proposition 1(d)} \\ &= w(\theta \wedge \phi) + w(\eta \wedge \phi) \quad \text{by P2 for } w \end{aligned}$$

and dividing by $w(\phi)$ gives the result.

For P3, note that

$$(\exists x\, \psi(x) \wedge \phi) \equiv \exists x\, (\psi(x) \wedge \phi)$$

$$\left(\bigvee_{i=1}^{n} \psi(a_i)\right) \wedge \phi \equiv \bigvee_{i=1}^{n} (\psi(a_i) \wedge \phi)$$

so using Proposition 1(d) and P3 for w,

$$\begin{aligned} w(\exists x\, \psi(x) \wedge \phi) &= w(\exists x\, (\psi(x) \wedge \phi)) \\ &= \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} (\psi(a_i) \wedge \phi)\right) \\ &= \lim_{n\to\infty} w\left(\left(\bigvee_{i=1}^{n} \psi(a_i)\right) \wedge \phi\right) \end{aligned}$$

and the result follows after dividing both sides by $w(\phi)$. ■

In this proof we have quoted the properties of w coming from P1-3 and Proposition 1 that we were using at each step. From now on however these properties will generally be assumed without explicit mention.

Conditional probabilities appear frequently in this subject in large part because it is generally assumed that they model 'updating'. To explain this suppose that on the basis of some knowledge (possibly empty) about the structure s/he is inhabiting our agent has settled on w as his subjective assignment of probabilities. Now suppose that the agent learns that $\phi \in SL$ is also true in this world. How should the agent 'update' w? The 'received wisdom' is that the agent should replace w by $w(.\,|\,\phi)$ (tacitly assuming that the agent had anticipated this eventuality by arranging that $w(\phi) > 0$). In justification for this there is again a Dutch Book argument, but now allowing also conditional bets (see the exercise at the end of this section).

Accepting this interpretation of conditional probability has a major consequence for the motivating problem in Inductive Logic, to wit:

> $\mathcal{Q}$: *In the situation of zero knowledge, logically, or rationally, what belief should our agent give to a sentence $\theta \in SL$ being true in M?*

For if we instead consider an agent who *does* have some knowledge K of the form

$$\{\, \phi_1, \phi_2, \ldots, \phi_n \,\}$$

then we can imagine that this agent first chose his/her subjective probability function w on the basis of *zero* knowledge and then learnt K and updated w to $w(.\,|\,\bigwedge_{i=1}^{n} \phi_i)$, assuming of course that the agent gave non-zero probability to

$\bigwedge_{i=1}^{n} \phi_i$ in the first place. Accepting this scenario means that $\mathcal{Q}$ is really the key issue, at least for knowledge K of this form, and in part justifies our focusing on this question.[9]

One difficulty one faces when presenting principles centred around conditional probabilities is that the conditioning sentence, ϕ in the account above, may have zero probability, leading to the need to be constantly introducing provisos. A useful convention to avoid this, which we shall adopt throughout, is to agree that an expression such as

$$w(\theta \,|\, \phi) = c, \quad \text{or} \quad w(\theta \,|\, \phi) = w(\eta \,|\, \zeta),$$

is shorthand for

$$w(\theta \wedge \phi) = cw(\phi), \quad w(\theta \wedge \phi)w(\zeta) = w(\eta \wedge \zeta)w(\phi) \quad \text{respectively.}$$

Clearly if the denominator(s) are non-zero these amount to the same thing whilst if the denominator(s) is zero the expression still has meaning (and, interestingly, usually the meaning we would wish it to have).

Exercise Suppose that in our betting scenario introduced in the Dutch Book section we also allow *conditional bets* on, or against θ *given* ϕ. That is, for stake $s > 0$ and $0 \leq p \leq 1$ the agent is offered a choice of one of two wagers:

(CBet1$_p$) Win $s(1-p)$ if $M \vDash \theta \wedge \phi$, lose sp if $M \vDash \neg\theta \wedge \phi$,

(CBet2$_p$) Win sp if $M \vDash \neg\theta \wedge \phi$, lose $s(1-p)$ if $M \vDash \theta \wedge \phi$,

where all bets are off (so no one wins or loses anything) if it transpires that $M \models \neg\phi$.

Show that if we define $CBel(\theta \,|\, \phi)$ to be the supremum of the $p \in [0,1]$ such that CBet1$_p$ is acceptable to the agent then the 'no Dutch Book condition' forces that

$$CBel(\theta \,|\, \phi) = \frac{Bel(\theta \wedge \phi)}{Bel(\phi)}$$

where Bel was as defined already for non-conditional bets.

The Completely Independent Solution

At first sight it might seem that there is an obvious answer to question $\mathcal{Q}$. Namely if I know nothing then as far as I am concerned for any predicate R_j and constant a_i $R_j(a_i)$ is just as likely as $\neg R_j(a_i)$ so that I should give them both the same probability, perforce 1/2 by (P4). Again if I know nothing then *surely*

[9]The corresponding issue has been the subject of rather detailed investigation in the case of *Propositional Uncertain Reasoning*, even for the case of knowledge K which itself expresses uncertainties. We refer the reader to [23] (especially chapter 6).

I am not justified in assuming any dependencies between the distinct $\pm R_j(a_i)$, so I should treat them as statistically independent. In other words for distinct $R_{j_1}(a_{i_1}), R_{j_2}(a_{i_2}), \ldots, R_{j_m}(a_{i_m})$ we should set

$$w(\pm R_{j_1}(a_{i_1}) \wedge \pm R_{j_2}(a_{i_2}) \wedge \ldots \wedge \pm R_{j_m}(a_{i_m})) = (1/2)^m.$$

From earlier remarks it is clear that this w extends to a probability function on L which for obvious reasons we refer to as the *Completely Independent Solution*.[10]

Unfortunately, as an assignment of *subjective probabilities* this w suffers the widely held criticism that it *denies induction*. To explain this let us suppose that we have a unary predicate R, about which we initially know nothing, and we assign probabilities according to this w, so in particular

$$w(R(a_1) \wedge R(a_2) \wedge ... \wedge R(a_m)) = 2^{-m}.$$

We now learn that $R(a_n)$ holds for $n = 1, 2, \ldots, 100$. Most of us, I imagine, would be inclined, on the basis of this new evidence, to give a rather high probability to $R(a_{101})$ also holding. However, according to the argument above we should assign probability

$$w(R(a_{101}) \,|\, R(a_1) \wedge R(a_2) \wedge ... \wedge R(a_{100})) = 2^{-100}/2^{-101} = 1/2 \;\; !$$

In other words we would be giving $R(a_{101})$ the same probability as if we had never learnt that $R(a_n)$ holds for $n = 1, 2, \ldots, 100$. This additional information then would have no relevance, there would be a total lack of learning, or *induction*, here. This seems counter-intuitive and scarcely deserving of the epithet 'rational'.[11]

Symmetry and de Finetti's Theorem

In the foregoing discussion we have have frequently mentioned rational principles, that is principles we might entertain our agent adhering to because of their purported rationality, but so far we have not actually specified any such principles. We shall now put that to rights by mentioning three principles which are so basic, and reasonable, that we shall actually take them as standing assumptions in what follows.

As we shall see in the sections which follow most of the principles which have been proposed to date are justified by considerations of symmetry, relevance or irrelevance. Of these it is the principles based on symmetry (such as our original

[10] In fact it is a probability function that we shall meet twice more in the future, as c^L_∞ and w^0_L.

[11] But if there is to be induction, how much of it should there be, and how can it be explained? This is an issue we shall consider again later when we come to discuss the notion of *relevance.*

coin tossing example) which seem the most compelling, the argument being that if we can demonstrate a symmetry in the situation then it would be irrational of the agent to break that symmetry when assigning probabilities.

One obvious such symmetry relates to the constants $a_1, a_2, a_3, \ldots$. The agent has no reason to treat these any differently[12], a consideration which leads to:

The Constant Exchangeability Principle, Ex
For $\phi(x_1, x_2, \ldots, x_m) \in FL$ and any vectors $b_1, b_2, \ldots, b_m$ and $b'_1, b'_2, \ldots, b'_m$ of (distinct) constants,[13]

$$w(\phi(b_1, b_2, \ldots, b_m) = w(\phi(b'_1, b'_2, \ldots, b'_m)).$$

Equivalently, for any permutation $\sigma \in \mathsf{S}_{\mathbb{N}^+}$ (= the set of permutations of $\mathbb{N}^+$),

$$w(\phi(a_1, a_2, \ldots, a_m)) = w(\phi(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)})).$$

Although in this statement of Ex $\phi(b_1, b_2, \ldots, b_m)$ can be any sentence of L it is easy to see that it is enough to take it to be a state description. Thus Ex is equivalent to the assertion that

$$w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) \text{ } \textit{depends only the vector } \langle m_1, m_2, \ldots, m_{2^q}\rangle \text{ where } m_j = |\{i \,|\, h_i = j\}|. \tag{10}$$

In a similar fashion we might argue that as far as our zero knowledge agent is concerned the atoms simply determine a partition of the a_i and that there is no reason to treat any one atom differently from any other. This leads to:

The Atom Exchangeability Principle, Ax

For any permutation τ of $\{1, 2, \ldots, 2^q\}$ and constants $b_1, b_2, \ldots, b_m$,

$$w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) = w\left(\bigwedge_{i=1}^{m} \alpha_{\tau(h_i)}(b_i)\right). \tag{11}$$

Equivalently Ax asserts that the left hand side of (11) depend only on the *multiset* $\{m_1, m_2, \ldots, m_{2^q}\}$ where as in Ex $m_j = |\{i \,|\, h_i = j\}|$.

Obviously requiring w to satisfy either of these principles severely restricts the possible choices for w. We now describe a simple, but important, family of probability functions satisfying Ex.

[12]The subscripts on the a's are simply to allow us to them list them easily. The agent is not supposed to 'know' that a_1 comes before a_2 which comes before . . . on our list.

[13]Recall our convention that when we write a formula as $\phi(x_1, x_2, \ldots, x_m)$ then it is assumed to not mention any constants, unless otherwise stated.

Let

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

and let

$$\vec{c} = \langle c_1, c_2, \ldots, c_{2^q} \rangle \in \mathbb{D}_{2^q}.$$

Now define $w_{\vec{c}}$ on a state description

$$\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$$

by

$$w_{\vec{c}} \left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i) \right) = \prod_{i=1}^{m} c_{h_i} = \prod_{j=1}^{2^q} c_j^{m_j}$$

where $m_j = |\{ i \,|\, h_i = j \}|$ for $i = 1, 2, \ldots, 2^q$. It can be quite easily checked that $w_{\vec{c}}$ extends uniquely to a probability function on $QFSL$ satisfying P1-2 and then, again uniquely by Theorem 7, to a probability function on SL satisfying Ex.

It turns out that by the following celebrated representation theorem due to B. de Finetti these $w_{\vec{c}}$ are the building blocks of all the probability functions on L satisfying Ex.

de Finetti's Representation Theorem 10 *Let L be a unary language with q predicates and let w be a probability function on SL satisfying Ex. Then there is a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ such that*

$$\begin{aligned} w \left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i) \right) &= \int_{\mathbb{D}_{2^q}} \prod_{j=1}^{2^q} x_i^{m_j} \, d\mu(\vec{x}) \\ &= \int_{\mathbb{D}_{2^q}} w_{\vec{x}} \left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i) \right) d\mu(\vec{x}) \end{aligned} \tag{12}$$

where $m_j = |\{i \,|\, h_i = j\}|$ for $j = 1, 2, \ldots, 2^q$.

Conversely, given a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ the function w defined by (12) extends (uniquely) to a probability function on SL satisfying Ex.

Proof To simplify the notation assume that $q = 1$, the full case being an immediate generalization. So there are just two atoms, $\alpha_1(x) = P_1(x)$ and $\alpha_2(x) = \neg P_1(x)$. Let

$$w(n, k) = w \left(\bigwedge_{i=1}^{n+k} \alpha_{h_i}(b_i) \right) \tag{13}$$

where $n = |\{i \,|\, h_i = 1\}|$ and $k = |\{i \,|\, h_i = 2\}|$. Notice that since w satisfies Ex the order of the h_i and the choice of distinct constants $b_1, b_2, \ldots, b_{n+k}$ here is immaterial. Notice that for fixed n, k there are $\binom{n+k}{n}$ distinct possibilities for the ordering of the $h_1, h_2, \ldots, h_{n+k}$ here. Let $r > n + k$. Then from (9)

$$1 = w(\top) = \sum_{r_1+r_2=r} \binom{r}{r_1} w(r_1, r_2) \tag{14}$$

and

$$w(n, k) = \sum_{\substack{r_1+r_2=r \\ n \le r_1,\, k \le r_2}} \binom{r-n-k}{r_1-n} w(r_1, r_2). \tag{15}$$

From (14) let μ_r be the discrete measure on $\mathbb{D}_2$ which puts measure

$$\binom{r}{r_1} w(r_1, r_2)$$

on the point $\langle r_1/r, r_2/r\rangle \in \mathbb{D}_2$. Then from (15) we obtain that $w(n, k)$ equals

$$\sum_{\substack{r_1+r_2=r \\ n \le r_1,\, k \le r_2}} \binom{r-n-k}{r_1-n} \binom{r}{r_1}^{-1} \binom{r}{r_1} w(r_1, r_2)$$

$$= \sum_{\substack{r_1+r_2=r \\ n \le r_1,\, k \le r_2}} \frac{r_1(r_1-1)\cdots(r_1-n+1)r_2(r_2-1)\cdots(r_2-k+1)}{r(r-1)\cdots(r-n-k+1)} \mu_r(\{\langle r_1/r, r_2/r\rangle\}). \tag{16}$$

The terms

$$\frac{r_1(r_1-1)\cdots(r_1-n+1)r_2(r_2-1)\cdots(r_2-k+1)}{r(r-1)\cdots(r-n-k+1)}$$

here must be within

$$\left(\frac{r_1}{r}\right)^n \left(\frac{r_2}{r}\right)^k \left(1 - \frac{(1-r_1^{-1})\cdots(1-(n-1)r_1^{-1})(1-r_2^{-1})\cdots(1-(k-1)r_2^{-1})}{(1-r^{-1})\cdots(1-(n+k-1)r^{-1})}\right) \tag{17}$$

of $(r_1/r)^n(r_2/r)^k$. By considering separately the cases $n, k > 0$, $r_1, r_2 \ge \sqrt{r}$ and $n > 0, r_1 < \sqrt{r}$ and $n = 0, r_1 < \sqrt{r}$ we see that (17) tends to 0 as $r \to \infty$ uniformly in r_1, r_2. Hence from (14) and (15) $w(n, k)$ equals the limit as $r \to \infty$ of

$$\sum_{\substack{r_1+r_2=r \\ n \le r_1,\, k \le r_2}} \left(\frac{r_1}{r}\right)^n \left(\frac{r_2}{r}\right)^k \mu_r(\{\langle r_1/r, r_2/r\rangle\}). \tag{18}$$

In turn this equals the limit of the same expressions but summed simply over $0 \leq r_1, r_2, r_1 + r_2 = r$ since from (14) (or trivially if $n = 0$),

$$\sum_{\substack{r_1+r_2=r \\ r_1<n,\, k\leq r_2}} \left(\frac{r_1}{r}\right)^n \left(\frac{r_2}{r}\right)^k \mu_r(\{\langle r_1/r, r_2/r\rangle\}), \quad \text{etc.}$$

tends to zero as $r \to \infty$.

In other words,

$$w(n,k) = \lim_{r\to\infty} \int_{\mathbb{D}_2} x_1^n x_2^k \, d\mu_r(\langle x_1, x_2\rangle). \tag{19}$$

By Prokhovor's Theorem, see for example [8], since $\mathbb{D}_2$ is compact the μ_r have a subsequence μ_{i_r} weakly convergent to a countably additive measure μ, meaning that for any continuous function $f(x_1, x_2)$

$$\lim_{r\to\infty} \int_{\mathbb{D}_2} f(x_1, x_2) \, d\mu_{i_r}(x_1, x_2) = \int_{\mathbb{D}_2} f(x_1, x_2) \, d\mu(x_1, x_2).$$

Using this the required result follows from (19).

Finally the converse result, that functions w defined by (12) extend to probability functions on SL satisfying Ex is entirely straightforward. ■

From (12) it follows that the integrals

$$\int_{\mathbb{D}_2} f(x_1, x_2) \, d\mu(\langle x_1, x_2\rangle)$$

are uniquely determined for *any* polynomial $f(x_1, x_2)$, and hence μ must be the unique measure satisfying (12). We shall call this measure the *de Finetti prior of* w.

de Finetti's Representation Theorem has a number of valuable applications one of which will be given later when we come to consider 'relevance'.

Exercise
Describe probability functions $t_{\vec{c}}$ on L, for $\vec{c} \in \mathbb{D}_{2^q}$, such that any probability function w on L satisfying Ax is of the form

$$\int_{\mathbb{D}_{2^q}} t_{\vec{x}} \, d\mu(\vec{x})$$

for some measure μ on the Borel subsets of $\mathbb{D}_{2^q}$, and conversely.

Irrelevance and Carnap's Continuum

Before our next theorem it will be useful to introduce the probability functions c^L_λ on L. For $0 < \lambda \leq \infty$ the probability function c^L_λ on the unary language L ($= \{R_1, R_2, \ldots, R_q\}$ as usual) is defined by

$$c^L_\lambda(\alpha_j(b_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)) = \frac{m_j + \lambda 2^{-q}}{n + \lambda} \tag{20}$$

where $m_j = |\{h_i \mid h_i = j\}|$, the number of times the atom $\alpha_j(x)$ occurs amongst the $\alpha_{h_i}(x)$. [When $\lambda = \infty$ this right hand side is just taken to be 2^{-q}.]
For $\lambda = 0$ c^L_0 is defined by

$$c^L_0\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\right) = \begin{cases} 2^{-q} & \text{if } h_1 = h_2 = \ldots = h_n, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Notice that (20) does indeed determine the value of c^L_λ on all state descriptions, and in turn determine a unique probability function, since it gives that

$$\begin{aligned} c^L_\lambda\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\right) &= \prod_{j=1}^{n} c^L_\lambda(\alpha_{h_j}(b_j) \mid \bigwedge_{i=j+1}^{n} \alpha_{h_i}(b_i)) \\ &= \prod_{j=1}^{n} \left(\frac{r_j + \lambda 2^{-q}}{n - j + \lambda}\right) \\ &= \frac{\prod_{k=1}^{2^q} \prod_{j=0}^{m_k - 1}(j + \lambda 2^{-q})}{\prod_{j=0}^{n-1}(j + \lambda)} \end{aligned} \tag{22}$$

where r_j is the number of times that h_j occurs amongst $h_{j+1}, h_{j+2}, \ldots, h_n$ and m_k is the number of times k occurs amongst $h_1, h_2, \ldots, h_n$. Furthermore we see from (21) and (22) that the values of the c^L_λ are invariant under permutations of the constants and the atoms and hence the c^L_λ satisfy Ex and Ax.

The c^L_λ for $0 \leq \lambda \leq \infty$ are referred to as *Carnap's Continuum of Inductive Methods* and, for reasons that will become clear in the next section, have played a central role in Applied Inductive Logic.

In the very first example in this book, tossing a coin to decide ends at the start of a football match we already met the idea of disregarding 'irrelevant knowledge' in forming beliefs. Several principles have been proposed within Inductive Logic aimed at capturing this idea, or at least aspects of this idea. The most historically important of these, as we shall see shortly, was what became known (following a suggestion by I.J.Good) as Johnson's Sufficientness Principle, or Postulate. This

principle was a key assumption of Johnson in his ground breaking 1932 paper [17] and later in Carnap's development of Inductive Logic.

Johnson's Sufficientness Principle[14]

$$w(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)) \tag{23}$$

depends only n and $m = |\{\, i \mid h_i = j \,\}|$ i.e. the number of times that α_j occurs amongst the α_{h_i} for $i = 1, 2, \ldots, n$.

In other words, knowing (just) which atoms are satisfied by $a_1, a_2, \ldots, a_n$ the probability of a_{n+1} satisfying a particular atom α_j depends only on the sample size n and the number of these which have already satisfied α_j (and similarly for any distinct constants $b_1, b_2, \ldots, b_{n+1}$ by our standing assumption of Ex). Within the framework used by Carnap where we may have families of properties, such as colors, in place of these atoms this might be motivated by imagining we have an urn containing balls colored blue, green, red and yellow, say, and we are picking from this urn with replacement. Suppose we have made n previous picks and out of them m have been red. In that case is does indeed seem intuitively clear, that the probability assigned to the $n+1$st pick being red will only depend on n and m and not on the distribution of the colors blue, green and yellow amongst the remaining $n - m$ balls.

Of course this particular motivation for JSP would require our agent to first assume that the ambient world had been, or was being, decided by picking atoms from some urn. In the situation of 'zero knowledge' that seems quite an assumption, though of course one could certainly entertain JSP as a reasonable, rational, principle without making any such assumption about the world.

Our next theorem, proved original by Johnson in [17] and later by Kemeny and Carnap & Stegmüller [7], shows why JSP has held such an esteemed position in Inductive Logic. Before that however we prove a useful lemma.

Lemma 11 *JSP implies Ax.*

Proof Since

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \prod_{j=1}^{n} w(\alpha_{h_j}(a_j) \mid \bigwedge_{i=j+1}^{n} \alpha_{h_i}(a_i))$$

(with both sides zero if not all the conditional probabilities are defined) JSP gives that this right hand side is invariant under permutations of atoms. Hence so is the left hand side and this gives that w satisfies Ax. ■

[14] As usual we apply the convention agreed on page 17 in the case when this conditional probability is not defined because the denominator is zero.

Theorem 12 *Suppose the unary language L has at least two relations, i.e. $q \geq 2$. Then the probability function w on L satisfies JSP if and only if $w = c_\lambda^L$ for some $0 \leq \lambda \leq \infty$.*

Proof It is clear from their defining equations that the c_λ^L satisfy JSP.

For the other direction assume that w satisfies JSP. Then by the previous lemma w satisfies Ax, a property we will use henceforth without further mention. Thus

$$1 = w\left(\bigvee_{i=1}^{2^q} \alpha_i(a_1)\right) = \sum_{i=1}^{2^q} w(\alpha_i(a_1))$$

so

$$w(\alpha_i(a_1)) = 2^{-q}, \quad \text{for } i = 1, 2, \ldots, 2^q. \tag{24}$$

Now suppose that

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = 0$$

for some state description. We may assume that n here is minimal. By (24) $n > 1$. If, say, $h_1 = h_2$ then by de Finetti's Theorem

$$\int_{\mathbb{D}_{2^q}} x_{h_2}^2 \prod_{i=3}^{n} x_{h_i}\, d\mu(\vec{x}) = 0,$$

where μ is the de Finetti prior for w, so

$$w\left(\bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\right) = \int_{\mathbb{D}_{2^q}} x_{h_2} \prod_{i=3}^{n} x_{h_i}\, d\mu(\vec{x}) = 0$$

contradicting the minimality of n.

Hence all the h_i must be different, so by JSP

$$0 = w\left(\alpha_{h_1}(a_1) \mid \bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\right) = w\left(\alpha_1(a_1) \mid \bigwedge_{i=2}^{n} \alpha_2(a_i)\right)$$

and we must have

$$w\left(\alpha_1(a_1) \wedge \bigwedge_{i=2}^{n} \alpha_2(a_i)\right) = 0.$$

Given the previous conclusion there cannot be any repeated atoms in $\bigwedge_{i=2}^{n} \alpha_2(a_i)$ so we must have $n = 2$.

This means that

$$w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = 0$$

whenever the h_i are not all equal. So in view of (24) we must have that

$$w\left(\bigwedge_{i=1}^{m}\alpha_j(a_i)\right) = 2^{-q} \quad \text{for } m \geq 1,\ j = 1, 2, \ldots, 2^q,$$

in other words $w = c_0^L$.

So now assume that w is non-zero on all state descriptions and the conditional probabilities in JSP are well defined and non-zero. Since (23) depends only on n and m denote it by $g(m, n)$.
Then from (24),

$$g(0, 0) = 2^{-q}.$$

Similarly

$$1 = w\left(\bigvee_{i=1}^{2^q}\alpha_i(a_2) \,|\, \alpha_j(a_1)\right) = \sum_{i=1}^{2^q} w(\alpha_i(a_2) \,|\, \alpha_j(a_1)) = g(1, 1) + (2^q - 1)g(0, 1),$$

so

$$g(1, 1) + (2^q - 1)g(0, 1) = 1. \tag{25}$$

By de Finetti's Theorem and Ax,

$$\begin{aligned}\int_{\mathbb{D}_{2^q}} x_1^2\, d\mu(\vec{x}) &= \int_{\mathbb{D}_{2^q}} (x_1^2 + x_2^2)/2\, d\mu(\vec{x}) \\ &\geq \int_{\mathbb{D}_{2^q}} x_1 x_2\, d\mu(\vec{x})\end{aligned}$$

so

$$g(1, 1) \geq g(0, 1).$$

Also, from (25),

$$1 \geq g(1, 1) \geq 2^{-q}.$$

Hence for some $0 \leq \lambda \leq \infty$,

$$g(1, 1) = \frac{1 + 2^{-q}\lambda}{1 + \lambda}, \quad g(0, 1) = \frac{2^{-q}\lambda}{1 + \lambda}, \tag{26}$$

the second of these following from the first and (25). Indeed we must have $\lambda > 0$ since $g(0, 0) > 0$.

We now show by induction on $n \in \mathbb{N}$ that for this same λ

$$g(m, n) = \frac{m + \lambda 2^{-q}}{n + \lambda}, \tag{27}$$

which forces w to satisfy the equations defining c_λ^L.

We have already proved (26) for $n = 0, 1$ so assume that $n \geq 1$ and (27) holds for n. The idea is to use JSP to derive a suitable set of equations which force (27) to also hold for $n + 1$.

Firstly for $r + s = n + 1$, and distinct $1 \leq m, k \leq 2^q$ we have that

$$\begin{aligned} 1 &= w\left(\bigvee_{h=1}^{2^q} \alpha_h(a_{n+1}) \mid \bigwedge_{i=1}^{r} \alpha_m(a_i) \wedge \bigwedge_{i=r+1}^{n+1} \alpha_k(a_i)\right) \\ &= \sum_{h=1}^{2^q} w\left(\alpha_h(a_{n+1}) \mid \bigwedge_{i=1}^{r} \alpha_m(a_i) \wedge \bigwedge_{i=r+1}^{n+1} \alpha_k(a_i)\right) \\ &= g(r, n+1) + g(s, n+1) + (2^q - 2)g(0, n+1). \end{aligned} \tag{28}$$

Secondly, adopting the convenient notation of writing $\alpha_{h_1}\alpha_{h_2}\ldots\alpha_{h_n}$ etc. for

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

we have that for $r + s + t = n$ and distinct $1 \leq m, j, k \leq 2^q$ (these exist since $2^q \geq 4$),

$$\begin{aligned} w(\alpha_m \mid \alpha_j\alpha_m^r\alpha_j^s\alpha_k^t) \cdot w(\alpha_j \mid \alpha_m^r\alpha_j^s\alpha_k^t) &= w(\alpha_m\alpha_j \mid \alpha_m^r\alpha_j^s\alpha_k^t) \\ &= w(\alpha_j \mid \alpha_m\alpha_m^r\alpha_j^s\alpha_k^t) \cdot w(\alpha_m \mid \alpha_m^r\alpha_j^s\alpha_k^t). \end{aligned}$$

Hence

$$g(r, n+1)g(s, n) = g(s, n+1)g(r, n). \tag{29}$$

In particular taking $s = 0$ here and using the inductive hypothesis (27) for n gives,

$$g(r, n+1) = (r\lambda^{-1}2^q + 1)g(0, n+1). \tag{30}$$

Taking respectively $r = 1,\ s = n$ in (28) and substituting for $g(1, n+1), g(n, n+1)$ from (30) gives

$$(\lambda^{-1} + 2^q)g(0, n+1) + (n\lambda^{-1} + 2^q)g(0, n+1) = 1$$

and hence

$$g(0, n+1) = \frac{\lambda 2^{-q}}{n + 1 + \lambda}.$$

Substituting in (29) now gives (27) too for $n + 1$ and $r = 1, 2, \ldots, n$, and finally also for $r = n + 1$ by using (28) with $r = 0,\ s = n + 1$. ■

Relevance and beyond

In our initial example in the introduction, of forming beliefs about the outcome of a coin toss, we noted three important considerations: symmetry, irrelevance

and relevance. We have already considered to some extent the first two of these (though we shall have more to say later) and we now come to consider the third, relevance. In that example we observed that knowing that the previous eight coin tosses by this referee had all been heads would surely have seemed relevant as far as the imminent toss was concerned.

Such an example seems to be a rather typical source of 'relevance': Namely certain patterns in past occurrences suggest that a similar pattern will be likely to repeat in the future. For example if I am looking out of my window and of the 6 birds I see fly past all but the third are headed left to right I might reasonably expect that more than likely the next bird observed will also be heading that way, though precisely quantifying what I mean by 'more than likely' might still be far from obvious.

Such examples were very much the focus of interest, and application, in Carnap's early studies in Inductive Logic with little attention being paid to our problem of the assignment of beliefs, or probabilities, in general. In these examples Carnap et al assumed the position that such past evidence should be incorporated via conditioning (as explained already on page 16) so in its simplest form 'relevance' tells us that we should have, formalizing the above bird example say, that

$$w(R_1(a_7) \mid R_1(a_1) \wedge R_1(a_2) \wedge \neg R_1(a_3) \wedge R_1(a_4) \wedge R_1(a_5) \wedge R_1(a_6)) > 1/2. \quad (31)$$

A problem here of course is saying by how much the left hand side should exceed a half.

As something of an aside at this point it might have appeared, from the time that question $\mathbb{Q}$ was first posed, that it had an an obvious answer: Namely, if nothing is known then there is no reason to suppose that any $R_j(a_i)$ is any more probable that $\neg R_j(a_i)$ (so both get probability 1/2), nor any reason to suppose that these $R_j(a_i)$ are in any way (stochastically) dependent. It is easy to see that there is just one probability function, c^L_∞ in fact, whose assignment of values satisfy this property. Unfortunately it gives, for example, equality in (31) rather the desired inequality. So despite its naive appeal this (apparent) flaw means that the subject of Inductive Logic has not immediately accepted the supremacy of this probability function (and packed up and gone home as a result) but instead has blossomed into more esoteric considerations.

Of course the examples above in support of 'relevance' are based on particular interpretations, which as we have stressed would not be available to the zero knowledge agent. On the other hand the agent could certainly introspect along the lines that 'if I learnt $R_1(a_1) \wedge R_1(a_2) \wedge \neg R_1(a_3) \wedge R_1(a_4) \wedge R_1(a_5) \wedge R_1(a_6)$ then would it not be reasonable for me to entertain (31)?'.[15]

[15] To counter this however the agent could argue that according to the naively favored c^L_∞ such biased conditioning evidence is rather unlikely in the first place so such a thought experiment should not cause one to particularly alter one's views!

Returning however to the intuitive desirability of (31) and its ilk Carnap proposed a principle which, with some simplifications, can be formulated as:

The Principle of Instantial Relevance, PIR
For $\theta(x_1, x_2, \ldots, x_n) \in QFFL$ and atom $\alpha(x)$ of L,

$$w(\alpha(a_{n+2}) \,|\, \alpha(a_{n+1}) \wedge \theta(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \,|\, \theta(a_1, a_2, \ldots, a_n)). \quad (32)$$

This principle then captures the informal idea that in the presence of 'evidence' $\theta(a_1, a_2, \ldots, a_n)$ the observation that a_{n+1} satisfies $\alpha(x)$ should enhance one's belief that a_{n+2} will also satisfy $\alpha(x)$.[16]

It is one of the pleasing successes of this subject that in fact PIR is a consequence simply of Ex, a result first proved by Gaifman [13] (see also [16]).

Theorem 13 *Ex implies PIR*

Proof Let the probability function w on L satisfy Ex. Using the notation of (32) let $\alpha(x) = \alpha_1(x)$ and let $\theta(\vec{a})$ be logically equivalent to the disjunction of state descriptions

$$\bigvee_{k=1}^{r} \bigwedge_{i=1}^{n} \alpha_{h_{ik}}(a_i).$$

Then for μ the de Finetti prior for w,

$$w(\theta(\vec{a})) = \int_{\mathbb{D}_{2^q}} \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}) = A \text{ say},$$

$$w(\alpha_1(a_{n+1}) \wedge \theta(\vec{a})) = \int_{\mathbb{D}_{2^q}} x_1 \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}),$$

$$w(\alpha_1(a_{n+2}) \wedge \alpha_1(a_{n+1}) \wedge \theta(\vec{a})) = \int_{\mathbb{D}_{2^q}} x_1^2 \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x})$$

and (32) amounts to

$$\left(\int_{\mathbb{D}_{2^q}} x_1 \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}) \right)^2 \leq \left(\int_{\mathbb{D}_{2^q}} \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}) \right) \cdot \left(\int_{\mathbb{D}_{2^q}} x_1^2 \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}) \right). \quad (33)$$

[16] Of course since we are assuming Ex the particular constants a_i used here are irrelevant.

If $A = 0$ then this clearly holds (because the other two integrals are less or equal to A and greater equal zero) so assume that $A \neq 0$. In that case multiplying out show that (33) is equivalent to

$$\int_{\mathbb{D}_{2^q}} \left(x_1 A - \int_{\mathbb{D}_{2^q}} x_1 \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}) \right)^2 \sum_{k=1}^{r} \prod_{i=1}^{n} x_{h_{ik}} \, d\mu(\vec{x}) \geq 0. \qquad (34)$$

But obviously, being an integral of a non-negative function, (34) holds, as required. ■

Clearly we can obtain more from this proof. For we can only have equality in (33) if $A = w(\theta(\vec{a})) = 0$ or if $w(\theta(\vec{a})) \neq 0$ and x_1 is constant on a measure 1 set of μ.It is also interesting to note that we do not need to take an atom $\alpha(x)$ in the statement of PIR, a minor revision of the proof gives that

The Extended Principle of Instantial Relevance, EPIR
For $\theta(x_1, x_2, \ldots, x_n), \phi(x_1) \in QFFL$,

$$w(\phi(a_{n+2}) \,|\, \phi(a_{n+1}) \wedge \theta(a_1, a_2, \ldots, a_n)) \geq w(\phi(a_{n+2}) \,|\, \theta(a_1, a_2, \ldots, a_n)). \qquad (35)$$

An explanation, suggested by Zabell, as to *why* Theorem 13 holds is that by saying that the constants are exchangeable Ex is implying that the future should look like the past. In other words observed past tendencies should be expected to repeat in the future. Looked at from this angle then Ex seems to be a much more potent force than it might initially have appeared.

Up to now we have looked at manifestations of relevance coming from Ex. But what about when we assume Ax, or JSP, might we not get similar, and stronger, such relevance principles? The answer is yes, but at this time the precise situation is still unclear, see [28] and [22].

Another Continuum of Inductive Methods

We have seen in the previous sections that if w satisfies Ex + JSP then w is a member of Carnap's Continuum of Inductive Methods, i.e. $w = c_\lambda^L$ for some $0 \leq \lambda \leq \infty$. Since Ex and JSP appear such natural and attractive principles it would seem unlikely that there could be other appealing principles around which would lead to an essentially different family of probability functions. However as we shall now see that is in fact the case.

Informally the Extended Principle of Instantial Relevance tells us that in the presence of information, or knowledge, $\psi(a_1, a_2, \ldots, a_n)$, learning that $\theta(a_{n+1})$ holds should enhance our belief that $\theta(a_{n+2})$ will hold. But what if instead of learning $\theta(a_{n+1})$ we had learnt only some consequence $\phi(a_{n+1})$ of $\theta(a_{n+1})$, should

this not also act as positive support for $\theta(a_{n+2})$ holding and so also enhance, or at least not diminish, belief in $\theta(a_{n+2})$? This intuition is summed up in the following principle:

The Generalized Principle of Instantial Relevance, GPIR
For $\theta(x_1), \phi(x_1), \psi(x_1, x_2, \ldots, x_n) \in QFFL$, if $\theta(x_1) \models \phi(x_1)$ then

$$w(\theta(a_{n+2}) \,|\, \phi(a_{n+1}) \wedge \psi(a_1, a_2, \ldots, a_n)) \geq w(\theta(a_{n+2}) \,|\, \psi(a_1, a_2, \ldots, a_n)). \quad (36)$$

At this point one might wonder if there should not also be such a principle for when $\phi(x_1) \models \theta(x_1)$. In fact such a principle would be equivalent to GPIR as we now show.

Suppose that $\phi(x_1) \models \theta(x_1)$ and let $\vec{a} = a_1, a_2, ..., a_n$. Notice that since

$$\begin{aligned} w(\theta(a_{n+2}) \,|\, \psi(\vec{a})) \;&=\; w(\theta(a_{n+2}) \,|\, \phi(a_{n+1}) \wedge \psi(\vec{a})) \cdot w(\phi(a_{n+1}) \,|\, \psi(\vec{a})) \\ &\quad + w(\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1}) \wedge \psi(\vec{a})) \cdot w(\neg\phi(a_{n+1}) \,|\, \psi(\vec{a})) \end{aligned}$$

and

$$w(\phi(a_{n+1}) \,|\, \psi(\vec{a})) + w(\neg\phi(a_{n+1}) | \psi(\vec{a})) = 1,$$

$$\begin{aligned} & w(\theta(a_{n+2}) \,|\, \phi(a_{n+1}) \wedge \psi(\vec{a})) \geq w(\theta(a_{n+2}) \,|\, \psi(\vec{a})) \\ \iff \quad & w(\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1}) \wedge \psi(\vec{a})) \leq w(\theta(a_{n+2}) \,|\, \psi(\vec{a})) \\ \iff \quad & w(\neg\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1}) \wedge \psi(\vec{a})) \geq w(\neg\theta(a_{n+2}) \,|\, \psi(\vec{a})), \end{aligned}$$

from which the result follows since $\theta(x_1) \models \phi(x_1)$ just if $\neg\phi(x_1) \models \neg\theta(x_1)$.

A second desirable property of our rational choice probability function w that we shall need here is that of Regularity.

The Principle of Regularity, Reg
For $\theta(\vec{a}) \in QFSL$ non-contradictory, $w(\theta(\vec{a})) > 0$.

Clearly since $w(\theta)$ is the sum of the $w(\Phi(\vec{a}))$ for $\Phi(\vec{a})$ a state description logically implying $\theta(\vec{a})$ it is enough in this definition to limit $\theta(\vec{a})$ to a state description. In that form the principle seems entirely reasonable, after all given a state description

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

what justification could one possibly find, in the total absence of any knowledge, for giving it zero probability?[17]

[17]Having said that there an argument is presented in [25] that w should be c_0^L and c_0^L does not satisfy Reg, for example it gives zero probability to the above conjunction when not all the h_i are equal.

In contrast however one might think that there is a possible reason for giving certain quantified sentences probability zero, for example $\forall x\, P_1(x)$, on the grounds that 'anything that can happen should at some point happen'. Arguing informally then as our rational agent might there should always be *some* a_i such that $\neg P_1(a_i)$, and by this one instance falsifying $\forall x\, P_1(x)$. Of course one can also see the other side of this, the agent arguing that it should at least be *possible* that $\forall x\, P_1(x)$ holds and hence it should receive some non-zero probability, though maybe not much.

This problem of how much, if any, probability to assign to sentences of the form $\forall x\, \psi(x)$ for $\psi \in QFSL$ has been much debated by the more philosophical 'Inductive Logicians'. Proposals have been put forward , see for example [9], but they tend to appear rather ah hoc. What one would like (if one thinks that non-zero probability should be so assigned) is an arguably rational principle which forces this as a consequence. In my view we do not yet have such a principle.

Returning to the main matter of this section, combining the above desiderata we seem to be led to proposing that our rational choice of w on L should satisfy GPIR + Reg + Ex + Ax. In that case however we have reduced the possible choices of w to members of another continuum, as the next theorem indicates.

Theorem 14 *Let w satisfy Ex + Ax + Reg. Then w satisfies GPIR if and only if $w = w_L^\delta$ for some $-(2^q - 1)^{-1} \leq \delta < 1$ where*

$$w_L^\delta = 2^{-q} \sum_{i=1}^{2^q} w_{\vec{e_i}(\delta)} \tag{37}$$

and

$$\vec{e_i}(\delta) = \langle \gamma, \gamma, \ldots, \gamma, \gamma + \delta, \gamma, \ldots, \gamma \rangle \in \mathbb{D}_{2^q}$$

with the $\gamma + \delta$ in the ith place and (necessarily) $\gamma = 2^{-q}(1 - \delta)$.

In particular then

$$w_L^\delta \left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \right) = 2^{-q} \sum_{j=1}^{2^q} \gamma^{m - m_j} (\gamma + \delta)^{m_j}$$

where as usual $m_j = |\{i \,|\, h_i = j\}|$.

The proof of this theorem is, unfortunately, not short enough to include in this course, for the details see [21], and we shall content ourselves in the next section with a short comparison between the $c_{\vec{\lambda}}^L$ and these w_L^δ.

Comparing the c_λ^L and w_L^δ.

We first remark that it is usual to define w_L^δ even for $\delta = 1$ by (37), so now these probability functions are defined for all δ in the interval $[-(2^q-1)^{-1}, 1]$. When $\delta = 1$ we must have that $\gamma = 0$ and in this case it turns out that $w_L^1 = c_0^L$, so the two continua agree at this extreme point. They also agree when $\delta = 0$, $\lambda = \infty$, i.e. $w_L^0 = c_\infty^L$, but these are the only two points at which the continua intersect.

Both continua satisfy Ex and Ax and also Regularity except when $\delta = 1$ or $\delta = -(2^q-1)^{-1}$ or $\lambda = 0$. A further property they both satisfy (after a domain restriction for the w^δ) is a form of *language invariance*, a notion which we shall first spend a little time explaining since it turns out play an important, and rather natural, role in the modern development of the subject.

The pertinent consideration underlying the support for 'language invariance' is that is that whilst our agent may be trying to make a rational choice of probability function w on a language L we would not want to be tied to the requirement that L was all the language there is and ever could be. We would like it to be the case that if at some point the language L expanded to a language L^+ then w had an extension, w^+ say, to L^+, meaning that w^+ restricted to SL (denoted $w^+{\restriction}SL$, notice that $SL \subseteq SL^+$) equals w.

Indeed we might require more than w just having an extension w^+ to L^+. For if we are arguing that some property $\mathcal{P}$ was a rational requirement on w then surely we should require $\mathcal{P}$ to be a rational requirement on w^+ too. This leads to the following family of conditions or desiderata:

Language Invariance with $\mathcal{P}$
The probability function w on L satisfies *Language Invariance with* $\mathcal{P}$ if there is a family of probability functions $w^{(\mathcal{L})}$ on $\mathcal{L}$ for each finite unary relational language $\mathcal{L} \subset \{R_j \mid j \in \mathbb{N}^+\}$, each satisfying property $\mathcal{P}$, such that $w^{(L)} = w$ and $w^{(\mathcal{L}^+)}{\restriction}S\mathcal{L} = w^{(\mathcal{L})}$ whenever $\mathcal{L}, \mathcal{L}^+$ are such languages and $\mathcal{L} \subseteq \mathcal{L}^+$.

A particular example of this is when $\mathcal{P}$ is Ex + JSP. In this case for each $0 \leq \lambda \leq \infty$ the family of probability functions $c_\lambda^{\mathcal{L}}$, *for this fixed* λ, forms such a language invariant family satisfying Ex + JSP since it is straightforward (though messy) to check that if

$$\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i) \tag{38}$$

is a state description of L and $L^+ = L \cup \{R_{q+1}\}$ then

$$c_\lambda^L\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = c_\lambda^{L^+}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = \sum_{\Phi(\vec{a})} c_\lambda^{L^+}(\Phi(\vec{a}))$$

where the $\Phi(a_1, \ldots, a_n)$ range over the state descriptions for $a_1, \ldots, a_n$ in SL^+ which logically imply (38).

Another way to see this is to notice (by comparing the values both sides give to state descriptions) that the de Finetti's Representation Theorem for c^L_λ, where $0 < \lambda < \infty$ (the extremes $\lambda = 0, \infty$ are easy to check separately), gives

$$c^L_\lambda\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = \kappa^L \int_{\mathbb{D}_{2^q}} w^L_{\vec{x}}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) \prod_{i=1}^{2^q} x_i^{\lambda 2^{-q}-1}\, d\mu^L(\vec{x}) \qquad (39)$$

where κ^L is a normalizing factor and μ^L is the standard Lebesgue measure on $\mathbb{D}_{2^q} \subset \mathbb{R}^{2^q}$.

Hence if $m_j = |\{i \mid h_i = j\}|$ and we enumerate the atoms of L^+ in the order

$$\alpha_1(x) \wedge R_{q+1}(x),\ \alpha_1(x) \wedge \neg R_{q+1}(x),\ \alpha_2(x) \wedge R_{q+1}(x),\ \alpha_2(x) \wedge \neg R_{q+1}(x), \ldots,$$

$$\ldots, \alpha_{2^q}(x) \wedge R_{q+1}(x),\ \alpha_{2^q}(x) \wedge \neg R_{q+1}(x)$$

then

$$\begin{aligned} c^{L^+}_\lambda\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) &= \kappa^{L+} \int_{\mathbb{D}_{2^{q+1}}} w^{L^+}_{\vec{y}}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) \prod_{i=1}^{2^{q+1}} y_i^{\lambda 2^{-q-1}-1}\, d\mu^{L^+}(\vec{y}) \\ &= \kappa^{L+} \int_{\mathbb{D}_{2^{q+1}}} \prod_{i=1}^{2^q} (y_{2i-1} + y_{2i})^{m_i} \prod_{i=1}^{2^{q+1}} y_i^{\lambda 2^{-q-1}-1}\, d\mu^{L^+}(\vec{y}) \end{aligned}$$

and integrating over the y_{2i-1}, y_{2i} with $x_i = y_{2i-1} + y_{2i}$ for $i = 1, 2, \ldots, 2^q$ gives the right hand side of (39).

As something of an aside here the identity (39) suggests a possible criterion for choosing the λ in c^L_λ, a question to which Carnap devoted much thought. Namely, if we set $\lambda = 2^q$ then the factor $\prod_{i=1}^{2^q} x_i^{\lambda 2^{-q}-1}$ becomes just 1 so in this integral all the $w^L_{\vec{x}}$ are taken as being 'equally likely' in the mixture which produces c^L_λ. Indeed this choice has been proposed in the literature, it is usually referred to as the *straight rule* since in (20) it gives

$$c^L_\lambda(\alpha_j(b_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)) = \frac{m_j + 1}{n + 2^q}.$$

Unfortunately this choice of λ clearly does not preserve this property once we extend the language, or marginalize to a smaller language. Thus, paradoxically, the argument *for* the straight rule for one language is an argument *against* the straight rule at any other language, at least if one wishes to adopt a language invariant family of probability functions.

Turning now to the w_L^δ they also satisfy a version of language invariance, *provided we restrict* δ *to* $[0,1]$, this time with their defining property GPIR. And indeed the pattern is rather similar, if we extend L to L^+ by adding a new predicate symbol R_{q+1} then the required extension of w_L^δ to SL^+ is $w_{L^+}^\delta$, so the δ has remained the same but the γ have gone down to $\gamma/2$, necessarily since there are now twice as many coordinates.

Exercise
Find the de Finetti prior for w_L^δ when $0 \leq \delta \leq \infty$.

Having pointed out some similarities between the c_λ^L and the w_L^δ, where henceforth we restrict the δ to $[0,1]$, we shall now remark upon a couple of difference (of which there are rather many).

Firstly when we introduced JSP the argument for it was based on a consideration of *irrelevance*. At that point however you might well have though that that was hardly the most basic and intuitively attractive formulation of irrelevance and that instead the following principle was even more evident:

The Weak Irrelevance Principle, WIR
If $\theta, \phi \in QFSL$ and have no predicate symbols nor constants in common then $w(\theta \,|\, \phi) = w(\theta)$, equivalently $w(\theta \wedge \phi) = w(\theta) \cdot w(\phi)$.

In other words if θ and ϕ share no common language then learning ϕ should not, rationally, change the belief that our agent assigns to θ.

Given that the defining property of the c_λ^L, JSP, was itself based on considerations of irrelevance it may come as something of a surprise to find that for $0 < \lambda < \infty$ these probability functions do not satisfy WIR. Counter-examples abound, for example it is easy to check the arithmetic to confirm that for $\lambda \in (0, \infty)$,

$$c_\lambda^L(P_2(a_3) \wedge P_2(a_4) \,|\, P_1(a_1) \wedge P_1(a_2)) > c_\lambda^L(P_2(a_3) \wedge P_2(a_4)). \tag{40}$$

Notice that here $P_1(a_1) \wedge P_1(a_2)$ and $P_2(a_3) \wedge P_2(a_4)$ come from completely disjoint languages.

On the other hand the w_L^δ do satisfy WIR, see [20] (though they are not the only such probability functions, see [26] for a characterization).

Actually the failure of WIR as instanced in (40) can be viewed as a positive feature of the c_λ^L, a demonstration of their capacity to detect 'higher order' instantial relevance. For the relevant feature in (40) is that because both a_1 and a_2 satisfy $P_1(x)$, rather than one satisfying $P_1(x)$ and the other satisfying $\neg P_1(x)$, this enhances the probability that the constants a_3, a_4 have similar properties, and so are more likely to both satisfy $P_2(x)$ (or both satisfy $\neg P_2(x)$) than to differ on $P_2(x)$.

A second principle which divides the continua is:

Reichenbach's Axiom, RA
Let $\alpha_{h_i}(x)$ for $i = 1, 2, 3, ...$ be an infinite sequence of atoms of L. Then for $\alpha_j(x)$ an atom of L,

$$\lim_{n\to\infty}\left(w(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^{n}\alpha_{h_i}(a_i)) - \frac{u(n)}{n}\right) = 0$$

where $u(n) = |\{i \mid 1 \leq i \leq n \text{ and } h_i = j\}|$.

So this principle, which was attributed to Reichenbach after a suggestion by Putnam, see [6, p120], asserts that as knowledge of the atoms satisfied by the $a_1, a_2, \ldots, a_n, \ldots$ grow so w should treat this information like a *statistical sample* and give a value for the probability that the next, $n+1$st, case revealed will be $\alpha_j(a_{n+1})$ which gets arbitrarily close to the frequency of past instances of $\alpha_j(a_i)$.

If our rational agent is something of a statistician this would surely seem like a common sense advice. However this principle says nothing about the ultimate convergence of the $u(n)/n$, an assumption which seems to be somehow implicit in any statistical viewpoint our agent might take.

Whether or not this is truly a desirable principle it is easy to see that it holds for c^L_λ when $0 < \lambda < \infty$ since in this case

$$c^L_\lambda\left(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^{n}\alpha_{h_i}(a_i)\right) - \frac{u(n)}{n} = \frac{u(n) + 2^{-q}\lambda}{n+\lambda} - \frac{u(n)}{n}$$

$$= \frac{2^{-q}\lambda}{n+\lambda} - \frac{\lambda u(n)}{n(n+\lambda)} \to 0 \quad \text{as } n \to \infty.$$

On the other hand for $0 < \delta < 1$ (in fact also for $0, 1$) w^δ_L fails to satisfy RA. To see this consider the sequence of atoms $\alpha_{h_i}(x)$ where $h_i = 1$ if $i = 0 \bmod 3$ and $h_i = 2$ otherwise. Then with $j = 1$ and the above notation $u(n)/n \to 1/3$ as $n \to \infty$. However

$$w^\delta_L\left(\alpha_1(a_{3n}) \mid \bigwedge_{i=1}^{3n-1}\alpha_{h_i}(a_i)\right) = \frac{\gamma(\gamma+\delta)^{2n} + \gamma^{n+1}(\gamma+\delta)^n + (2^q-2)\gamma^{2n+1}}{(\gamma+\delta)^{2n} + \gamma^{n+1}(\gamma+\delta)^{n-1} + (2^q-2)\gamma^{2n}}$$

which tends to γ as $n \to \infty$, so RA fails.[18]

A final property according to which these continua differ is:

Recoverability.

[18]Of course if $\gamma = 1/3$ we can just change the frequency of the α_1.

A probability function w on L is *Recoverable* if whenever $\phi(a_1, a_2, \ldots, a_n)$ is a state description then there is another state description $\phi'(a_{n+1}, a_{n+2}, \ldots, a_h)$ such that $w(\phi \wedge \phi') \neq 0$ and for any quantifier free sentence $\theta(a_{h+1}, a_{h+2}, \ldots, a_{h+g})$,

$$w(\theta(a_{h+1}, a_{h+2}, \ldots, a_{h+g}) \mid \phi \wedge \phi') = w(\theta(a_{h+1}, a_{h+2}, \ldots, a_{h+g})).$$

In other words w is Recoverable if given any 'past history' as such a state description ϕ there is a possible 'future' state description ϕ' which will take us right back to where we started, at least as far as the quantifier free properties of the currently unobserved constants $a_{h+1}, a_{h+2}, \ldots$ are concerned.

The w_L^δ satisfy Recovery provided $0 \leq \delta < 1$, indeed a particular such w_L^δ is characterized by satisfying Li with Ex + Ax and *a single instance* of recovery. Clearly then the c_λ^L can never satisfy recovery for $0 < \lambda \leq \infty$, even a single instance of it. Details of these results may be found in [28].

I would not wish to give the impression that I was advocating the w_L^δ (for $\delta \in [0, 1]$) as 'the rational choice' any more (in fact rather less) than I would wish to push the c_λ^L, though they do have some interesting properties. Rather what is interesting here is that we can have two 'continua of inductive methods' both arising from ostensibly reasonable rationality requirements. (It is also interesting how many of such principles at least one of these continua satisfy.) In particular it is surprising how the apparently innocuous step up from PIR to GPIR has such a profound effect.

Conclusion

In this short course I have endeavored to present you with some of the main results to date for *Unary Inductive Logic*, that is where all the relation symbols are unary. The studies of Johnson and Carnap never got beyond the unary though Carnap certainly was aware that moving up to *Polyadic Inductive Logic*, where we allow relation symbols of any finite arity, was a future challenge. Over the last decade in particular there have been some serious moves in that direction, see for example [18], [19], and to all intents the area now seems ripe for discovery.

Appendix

Goodman's Grue Paradox originated in [14] and has since spawned a considerable literature and range of formulations, see for example [27].
We shall give here a pared down mathematician's version. Let *grue* stand for 'green before the 1st of next month, blue after'. Now consider the following statements:

All the emeralds I have ever seen have been green, so I should give high probability that any emerald I see next month will be green.

All the emeralds I have ever seen have been grue, so I should give high probability that any emerald I see next month will be grue.

The conclusion that advocates of this 'paradox' would have us conclude is that Carnap's hope of determining such probabilities by purely logical or rational considerations cannot succeed. For here are 'isomorphic' premises with different (contradictory even) conclusions so the conclusion cannot simply be a logical function of the available information.

As mathematicians we would counter that the key requirement that *all* the available information be made explicit is not being observed here. Indeed it is precisely such undisclosed knowledge that is causing us to even conclude that these two statements are contradictory. Namely the knowledge that an emerald cannot be both green and blue. Were we to substitute 'expensive' for 'blue' there would be no contradiction.

Carnap himself in his writings emphasized the requirement that for the conclusion to be a 'logical consequence' of the knowledge all the knowledge should be made explicit (in modern day terms that such reasoning is *non-monotonic*). However what the 'paradox' does show is the general impracticality of an Inductive Logic in the form that Carnap was proposing. For how in a real world situation could one hope to put *all* one's knowledge explicitly into the calculation? It is this failure which spelled the end of the programme as *applied* logic for most philosophers, though not, for us mathematical logicians, as a branch of *pure* logic.

References

[1] Ash, R.B., *Probability and Measure Theory,* 2nd edition, Academic Press, 1999, ISBN0120652021.

[2] Bartle, R.G., *Elements of Integration and Legesgue Measure*, Wiley Interscience, 1995.

[3] Carnap, R., *Logical Foundations of Probability*, University of Chicago Press, Chicago, Routledge & Kegan Paul Ltd., 1950.

[4] Carnap, R., *The continuum of inductive methods,* University of Chicago Press, 1952.

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

[6] 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.

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

[8] Dudley, R.M., *Real Analysis and Probability*, Chapman & Hall, 1989, ISBN 0-412-05161-3

[9] Earman, J., *Bayes or Bust?*, MIT Press, 1992.

[10] de Finetti, B., Sul significato soggettivo della probabilità, *Fundamenta Mathematicae,* 1931, **17**:298-329.

[11] de Finetti, B., *Theory of Probability, Volume 1*, Wiley, New York, 1974.

[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, Berkeley and Los Angeles, 1971, pp235-251.

[14] Goodman, N., A query on confirmation, *Journal of Philosophy*, 1946, **43**:383-385.

[15] Goodman, N., On infirmities in confirmation-theory, *Philosophy and Phenomenological Research*, 1947, **8**:149-151.

[16] 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.

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

[18] Landes, J., Paris, J.B. & Vencovská, Some aspects of polyadic inductive logic,*Studia Logica*, 2008, **90**:3-16.

[19] Landes, J., Paris, J.B. & Vencovská, A survey of some recent results on Spectrum Exchangeability in Polyadic Inductive Logic, submitted to *Knowledge, Rationality and Action.*

[20] Nix, C.J., *Probabilistic Induction in the Predicate Calculus*, Ph.D. Thesis, University of Manchester, 2005.

[21] 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.

[22] Paris, J.B., An observation on Carnap's Continuum and Stochastic Independencies, submitted to *Erkenntnis.*

[23] Paris, J.B. *The Uncertain Reasoner's Companion,* Cambridge University Press, 1994.

[24] Paris, J.B. & Vencovská, In defense of the Maximum Entropy Inference Process, *International Journal of Approximate Reasoning*, **17**(1):77-103, 1997.

[25] Paris, J.B. & Vencovská, A., Symmetry's End?, to appear in *Erkenntnis.*

[26] Paris, J.B. & Vencovská, A., A Note on Irrelevance in Inductive Logic, submitted to the *International Journal of Approximate Reasoning.*

[27] Stalker, D. ed.,*Grue! The New Riddle of Induction*, Open Court, 1994.

[28] Paris, J.B. & Waterhouse, P., Atom Exchangeability and Instantial Relevance, *Journal of Philosophical Logic*, 2009, **38**(3):313-332.
DOI: 10.1007/s10992-008-9093-3

[29] Williamson, J., *In Defence of Objective Bayesianism*, Oxford University Press, 2010.