Logique & Analyse **228** (2014), 489–511

PRINCIPLES OF REMEMBERING AND FORGETTING

Elizabeth Howarth* and Jeff Paris

Abstract

We propose two principles of inductive reasoning related to how observed information is handled by conditioning, and justify why they may be said to represent aspects of rational reasoning. A partial classification is given of the probability functions which satisfy these principles.

Keywords

Inductive Logic, Logical Probability, Rationality, Uncertain Reasoning.

Introduction

Each time we consider the question of how likely some event is to occur (or have occurred), we surely go through some process of weighing any relevant information we possess to assist us in reaching a conclusion. In the absence of any relevant experience we may, for example, employ symmetry principles to try to establish the number of apparently equally likely cases and reason from there. However, where experiential data are available we would surely want to examine these to help inform our prediction.

For example, presented with an untested coin I would assign a probability of one half to its landing heads, based on symmetry and supported by observations of previous tosses of similar coins. However, if a sequence of tosses of this particular coin provides a frequency very different from one half, I would probably reconsider my assignment at some stage and adjust it to correspond more closely with my observations.

The use of probability functions and conditioning to model belief and learning are important techniques in Inductive Logic, (see [5] or [18] for discussion and justification of this approach). There have been several principles proposed in Inductive Logic which are intended to capture some aspect of this idea that the probabilities one assigns should be informed in some way by one's experiences. For example, Carnap's Principle of Instantial

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

doi: 10.2143/LEA.228.0.3078172.

Relevance [2, Chapter 13], Reichenbach's Axiom, see [3, p. 20], and the Unary Principle of Induction (see [18, Chapter 20] for this and a host of related principles). The first and third of these express the notion that the more times one has seen something in the past, the more likely one is to see it in the future, while the second asserts more strongly that the probability one assigns to an event should shadow its observed frequency (whether or not this converges to a single value). Whichever of these or other formulations is preferred, it seems to be widely accepted that it is rational to be prepared to alter the probabilities one assigns in light of acquired knowledge or observations.

This raises the question of how fine such adjustment should be. In other words, how different should two experiences (i.e. sequences of observations[1]) need to be before they result in different probability functions? Any choice of a numerical measure of 'difference' would seem to be rather arbitrary, but such a choice may be avoided by requiring that *any* difference in experience should result in a different probability function.

This idea forms the basis of the Elephant Principle, named after the saying that 'an elephant never forgets'. This requires that a probability function should, after conditioning on different past observations, result in different predictions for future observations. This ensures that all learning is 'remembered' by being uniquely incorporated into the resulting assignment.

On the other hand, it would seem unreasonable that two sequences of observations which are essentially very similar could result in wildly different probabilities being assigned to possible future observations. Rationality would seem to require us to keep our adjustments proportionate somehow, which idea forms the basis of the Perspective Principle. Here also, an arbitrary numerical measure of 'likeness' or 'difference' between two experiences is avoided, using the requirement that, whenever different finite experiences are each followed by identical and arbitrarily long sequences of observations, the difference between the two resulting conditional probability functions concerning possible future observations, *eventually* becomes arbitrarily small.

In fact the standing assumption of Constant Exchangeability, introduced in the following section, ensures that the order of observations is irrelevant. Therefore, the somewhat far-fetched scenario described is actually just one example of the more general situation where two sequences of observations of equal but arbitrary length eventually contain so many matched pairs of outcomes compared to unmatched ones that the resulting conditional probability functions should, according to the Perspective Principle, become arbitrarily similar.

[1] We shall make clear precisely what we mean by an 'observation' in the next section.

While the Perspective Principle was developed as a counterbalance to the Elephant Principle, it may also be considered in its own right without reference to the latter. And, by the above arguments, both may be considered as requirements of *rationality*.

Context and Notation

The context here is the one common to a number of recent accounts of *unary* Pure Inductive Logic by the authors et al., see for example [16], [18]. Thus we assume that we are working in a first order language L with finitely many unary predicate symbols, P_1, P_2, ..., P_q say, countably many constants $a_1, a_2, a_3, \ldots$ and no equality or other relation, function or constant symbols. The intention here is that these constants a_i exhaust the universe, that is between them they name all individuals in the universe. Let SL and $QFSL$ denote respectively the set of sentences and the set of quantifier-free sentences of L and let $\mathcal{T}$ denote the set of structures for L with universe $\{a_i \mid i \in \mathbb{N}^+\}$ (and each constant symbol a_i interpreted as a_i).

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

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

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

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

where $\models$ is the semantic consequence relation for the logic of L.

We now ask how a supposedly rational agent inhabiting a structure in $\mathcal{T}$, but having no prior knowledge concerning which such structure, should assign probabilities $w(\theta)$ to the sentences $\theta \in SL$. Or putting it another way, to what extent does the requirement of rationality limit the agent's choice of probability function?

A standard procedure for investigating this question is to propose purportedly rational principles which one may feel the agent should observe and then investigate their consequences, typically, as in this paper, by characterizing the probability functions which satisfy them. Amongst such principles are some which seem so reasonable that they are frequently taken as given.

The first of these principles which we shall assume herein is:

Constant Exchangeability, Ex

If σ is a permutation of $1, 2, \ldots$ and $\theta(a_1, \ldots, a_n) \in SL$ mentions at most the constants $a_1, \ldots, a_n$ then $w(\theta(a_{\sigma(1)}, \ldots, a_{\sigma(n)})) = w(\theta(a_1, \ldots, a_n))$.

The argument for this principle is that there is complete symmetry between the constants, and hence between $\theta(a_1,\ldots,a_n)$ and $\theta(a_{\sigma(1)},\ldots,a_{\sigma(n)})$, so it would be irrational to assign these two sentences different probabilities. All the probability functions we shall consider will be assumed to satisfy Ex.

A second symmetry based principle which we shall be assuming later requires us to first introduce some notation.

By the *atoms*[2] $\alpha_1(x),\alpha_2(x),\ldots,\alpha_{2^q}(x)$ of L we mean the 2^q formulae obtained by going through all combinations of $\varepsilon_1,\ldots,\varepsilon_q\in\{0,1\}$ in

$$P_1(x)^{\varepsilon_1}\wedge P_2(x)^{\varepsilon_2}\wedge\ldots\wedge P_q(x)^{\varepsilon_q}$$

where for a sentence θ, $\theta^1=\theta, \theta^0=\neg\theta$.

Then, based on the idea that, in terms of assigned probability, there is no reason to treat one atom any differently from any other, we have the principle of:

Atom Exchangeability, Ax

For τ a permutation of $\{1,2,\ldots,2^q\}$,

$$w\left(\bigwedge_{r=1}^{m}\alpha_{g_r}(a_{s_r})\right)=w\left(\bigwedge_{r=1}^{m}\alpha_{\tau(g_r)}(a_{s_r})\right).$$

A *state description* for (distinct) $a_{s_1},a_{s_2},\ldots,a_{s_m}$ is a quantifier free sentence $\Theta(a_{s_1},a_{s_2},\ldots,a_{s_m})$ of the form

$$\bigwedge_{r=1}^{m}\alpha_{g_r}(a_{s_r}).$$

As here, upper case Greek letters will always be used to denote state descriptions. Regarding the discussion in the Introduction, we shall identify state descriptions with ‘observations’ which our agent may make, or more reasonably given his/her situation, *imagine* making.

It follows immediately from Ex that the probability assigned to a state description Θ depends only on its *signature*: $\langle m_1,\ldots,m_{2^q}\rangle$ where m_i = $|\{r\mid g_r=i\}|$, the number of times that atom α_i features in Θ, regardless of which constants instantiate which atoms. Therefore, the alternative notation

$$\bigwedge_{i=1}^{2^q}\alpha_i^{m_i}$$

[2] Not to be confused with ‘atomic formulae’, which for this language would be the formulae $P_j(x_i)$. Our ‘atoms’ correspond to what Carnap et al. dubbed ‘molecular Q-predicates’, this alternative nomenclature the result of us arriving here via a different route (Nonmonotonic Logic).

may be used for state descriptions when the (distinct) instantiating constants are sufficiently clear from the context.

By the Disjunctive Normal Form Theorem every $\phi(a_{s_1}, a_{s_2}, \ldots, a_{s_m}) \in QFSL$ is logically equivalent to a disjunction of (necessarily pairwise disjoint) state descriptions, from which we can show that the probability of ϕ is the sum of the probabilities of these state descriptions. (For this and similar basic facts about probability functions used in this paper we refer the reader to [15, page 162] or [18, Proposition 3.1].) Indeed by Gaifman's Theorem [6], a probability function is completely determined on the whole of *SL*, not just on *QFSL*, by its values on state descriptions.

In what follows we shall assume that we have fixed some particular ordering, $\alpha_1(x), \alpha_2(x), \ldots, \alpha_{2^q}(x)$, of the atoms of *L*. With this in place we can define an important family of probability functions on *SL* as follows. Let

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

and for $\vec{c} \in \mathbb{D}_{2^q}$ define

$$w_{\vec{c}}\left(\bigwedge_{r=1}^{m} \alpha_{g_r}(a_{s_r})\right) = \prod_{r=1}^{m} c_{g_r} = \prod_{i=1}^{2^q} c_i^{m_i} \tag{1}$$

where, as above, $m_i = |\{r | g_r = i\}|$. Then $w_{\vec{c}}$ extends to a probability function on *SL*. Clearly $w_{\vec{c}}$ satisfies Ex, though not Ax unless all the c_i are equal, i.e. have value 2^{-q}.

The $w_{\vec{c}}$ are important in Inductive Logic because of the following Representation Theorem of de Finetti, see [5] or, in the notation of this paper, [18, Theorem 9.1], which we shall be using frequently in what follows.

Theorem 1. *Let w be a probability function on SL satisfying Ex. Then there is a measure*[3] μ *on the Borel*[4] *subsets of* $\mathbb{D}_{2^q}$ *such that*

$$w\left(\bigwedge_{r=1}^{m} \alpha_{g_r}(a_{s_r})\right) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\left(\bigwedge_{r=1}^{m} \alpha_{g_r}(a_{s_r})\right) d\mu(\vec{x}),$$

$$= \int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{m_i}\, d\mu(\vec{x}) \tag{2}$$

[3] All measures in this paper will be taken to be normalized and countably additive.

[4] In other words the closure under complement and countable unions of the open subsets of, in this case, the relativized topology on $\mathbb{D}_{2^q} \subseteq \mathbb{R}^{2^q}$. This is sufficient to ensure that the functions $\vec{x} \mapsto w_{\vec{x}}(\theta)$ are indeed integrable with respect to μ for $\theta \in SL$.

where $m_i = |\{r \,|\, g_r = i\}|$.

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

The measure μ is known as the *de Finetti prior* of the function w.

The two main results of this paper, Theorems 2 and 5, give respectively a characterization of the Elephant Principle in the presence of Ax and a partial characterization the Perspective Principle. The value of such results is twofold. Firstly they can help us to locate a principle within the general landscape of putatively rational principles, indeed this will be the case for the second of our theorems. Secondly they may enable us to quickly deduce whether or not a particular probability function satisfies the characterized principle, which may help to elucidate the extent to which this choice of function may be considered 'rational'.

We apply our characterization results in this way to two families of probability functions, namely Carnap's well known Continuum of Inductive Methods and the somewhat recent Nix-Paris Continuum. Each consists of a continuum of probability functions, characterized up to a real parameter as satisfying certain somewhat attractive rational principles: Johnson's Sufficiency Postulate in the case of Carnap's Continuum (for $q > 1$), see [10], [11] or [4], and the Generalized Principle of Instantial Relevance (plus Language Invariance), see [13], in the case of the Nix-Paris Continuum.

In more detail, Carnap's Continuum, for our specific language L with q predicates, consists of the probability functions c_λ^L for $0 \leq \lambda \leq \infty$ specified[5] for $\lambda > 0$ by

$$c_\lambda^L\left(\alpha_j(a_{m+1}) \,|\, \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = \frac{m_j + \lambda 2^{-q}}{m + \lambda}, \tag{3}$$

where $m_j = |\{i \,|\, h_i = j\}|$, the number of times the atom $\alpha_j(x)$ occurs amongst the $\alpha_{h_i}(x)$ and we identify $(2^{-q} \cdot \infty)/\infty$ with 2^{-q}, and for $\lambda = 0$ by

$$c_0^L\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = \begin{cases} 2^{-q} & \textit{if } h_1 = h_2 = \ldots = h_m, \\ 0 & \textit{otherwise.} \end{cases}$$

[5] It is straightforward to show that this determines the value of c_λ^L on every state description, hence on every quantifier free sentence, and finally on all of *SL* by Gaifman's Theorem [6].

Note that c_∞^L is just Carnap's probability function m^* which gives all the $P_j(a_i)$ probability 1/2 and treats them as stochastically independent – and so is totally devoid of any 'learning by induction'.

In terms of de Finetti's Theorem

$$c_\lambda^L = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\, d\mu(\vec{x}),$$

for $0 < \lambda < \infty$, where

$$d\mu(\vec{x}) = \kappa \prod_{i=1}^{2^q} x_i^{\lambda 2^{-q}-1} d\rho(\vec{x}), \tag{4}$$

ρ is Lebesgue measure and κ is a normalizing constant.

The Nix-Paris continuum for L is made up of the probability functions w_L^δ for $0 \le \delta \le 1$ given by

$$w_L^\delta = 2^{-q} \sum_{j=1}^{2^q} w_{\vec{e}_j(\delta)} \tag{5}$$

where $\vec{e}_j(\delta) = \langle \gamma, \ldots, \gamma, \gamma + \delta, \gamma, \ldots, \gamma \rangle \in \mathbb{D}_{2^q}$, with $\gamma + \delta$ in the jth position and, necessarily, $\gamma = 2^{-q}(1-\delta)$. This also covers the two remaining cases $\lambda = 0, \infty$ of Carnap's Continuum, since $c_0^L = w_L^1$ and $c_\infty^L = w_L^0$.

Clearly the de Finetti prior of w_L^δ is just the point measure which places measure 2^{-q} on each of the 2^q points $\vec{e}_j(\delta)$ (or measure 1 on the single point $\langle 2^{-q}, 2^{-q}, \ldots, 2^{-q} \rangle$ if $\delta = 0$).

These two continua agree at their end points, precisely $c_0^L = w_L^1$ and $c_\infty^L = w_L^0$, but not anywhere else. Both satisfy Ex and Ax, though in general they have rather different properties (for a comparison see [19]) as indeed we shall see in the case of the two principles investigated in this paper.

The Elephant Principle

In this section we define the Elephant Principle and present a representation theorem for probability functions which satisfy the Elephant Principle together with Ax. We apply this theorem to show that members of Carnap's Continuum of probability functions, c_λ^L, satisfy the Elephant Principle except at the endpoints $\lambda \in \{0,1\}$, while members of the Nix-Paris continuum of probability functions, w_L^δ, fail to satisfy it.

As discussed in the Introduction, the motivation for the Elephant Principle is the idea that the probabilities assigned by a rational agent to future events should reflect its observations of past events, which surely means that, in some cases, different observations should result in different assignments.

If this notion is taken to its extreme, to avoid any arbitrary measure of when two observations are 'sufficiently different', the resulting principle is that *any* difference in observations should result in *some* difference in assignments. It could also be argued that such perfect recall is 'rational' is based on the tenet that information is valuable and should never be discarded; that you cannot do better by knowing less.[6,7] We formalize this idea using conditional probabilities, as follows.

Given a probability function w on SL and $\phi \in SL$, we define the conditional probability function $w(.|\phi): SL \to [0,1]$ to be a function which satisfies

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

so $w(\cdot|\phi)$ is a probability function if $w(\phi) > 0$. In what follows we will identify, e.g.,

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

with

$$w(\theta \wedge \phi) \cdot w(\phi') = w(\theta' \wedge \phi') \cdot w(\phi), \tag{6}$$

which will hold automatically if either $w(\phi) = 0$ or $w(\phi') = 0$.

Now imagine our agent, who had initially adopted a probability function w, making (or imagining making) an observation Γ about $a_1, \dots, a_g$ and consequently conditioning on this evidence to form $w(\cdot|\Gamma(a_1, \dots, a_g))$. Because Γ is a state description the agent is now in no doubt about the properties of $a_1, \dots, a_g$, so we are really only concerned with how the agent's updated probability function $w(\cdot|\Gamma)$ assigns probabilities to state descriptions involving constants from $a_{g+1}, a_{g+2}, a_{g+3}, \dots$.

For this reason we define, for a given state description $\Gamma(a_1, \dots, a_g)$ and a probability function w on SL such that $w(\Gamma) > 0$, a function $w_{*\Gamma}$ on *state descriptions*[8] $\Theta(a_1, \dots, a_n)$ of L by

$$w_{*\Gamma}(\Theta(a_1, \dots, a_n)) = w(\Theta(a_{g+1}, \dots, a_{g+n}) \,|\, \Gamma(a_1, \dots, a_g)).$$

Notice that because of our standing assumption that w satisfies Ex, $w_{*\Gamma}$ will also satisfy Ex.

6 For an interesting justification of this see [8].

7 On the other hand there are several 'rational principles' in Inductive Logic which work on the basis of prescribing certain sorts of information 'irrelevant', for example Johnson's Sufficientness Principle, JSP. The fact that Carnap's Continuum satisfies both JSP and EP however shows that in this case they touch on differing forms of 'information'.

8 In fact, see [14], $w_{*\Gamma}$ extends to probability function on SL and continues to satisfy the identity

$$w_{*\Gamma}(\theta(a_1, \dots, a_n)) = w(\theta(a_{g+1}, \dots, a_{g+n}) \,|\, \Gamma(a_1, \dots, a_g))$$

even for θ simply a sentence of L. We will not need this however in what follows.

We now define the Elephant Principle to formalize the idea that $w_{*\Gamma}$ should reflect or 'remember' the information $\Gamma(a_1,\ldots,a_g)$, or putting it another way that $w_{*\Gamma}$ and w fix $\Gamma(a_1,\ldots,a_g)$ (up to the orders of the atoms and the constants).

The Elephant Principle, EP

For $\Gamma = \bigwedge_{i=1}^{2^q} \alpha_i^{g_i}$ and $\Gamma' = \bigwedge_{i=1}^{2^q} \alpha_i^{h_i}$ state descriptions of a unary language L, a probability function w on SL satisfies EP if

$$w_{*\Gamma} = w_{*\Gamma'} \Leftrightarrow g_i = h_i \quad \text{for } i = 1, 2, \ldots, 2^q.$$

So if w satisfies EP, then $w_{*\Gamma} = w_{*\Gamma'}$ just if Γ and Γ' have the same signature, so that any acquired information is uniquely reflected in the way in which the agent assigns probability to possible future observations.

We now give a representation theorem characterizing those probability functions on *SL* which satisfy Ax + EP, after introducing some notation.

Let $\mathbb{N}_n$ denote $\{1,2,\ldots,n\}$. For $S \subset \mathbb{N}_{2^q}$ let

$$N_S = \{\vec{x} \in \mathbb{D}_{2^q} \mid x_i = 0 \Leftrightarrow i \in S\},$$

and note that these N_S partition $\mathbb{D}_{2^q}$.

Theorem 2. *Suppose that w is a probability function satisfying Ax with de Finetti prior μ, and let $z = min\{|S| \text{ such that } \mu(N_S) > 0\}$. Then w fails EP just if there is some $X \in \mathbb{R}$ such that*

$$\mu(\{\vec{x} \in N_S \mid \prod_{i \notin S} x_i = X\}) = \mu(N_S),$$

for every $S \subset \mathbb{N}_{2^q}$ such that $|S| = z$.

In other words, if z is the size of the smallest $S \subset \mathbb{N}_{2^q}$ such that $\mu(N_S) > 0$, then w fails EP just if for every $S \subset \mathbb{N}_{2^q}$ of size z, all the measure in N_S is concentrated on those $\vec{x}$ for which the product of the coordinates, $\prod_{i=1}^{2^q} x_i$, equals some fixed X. It is not clear (at least to the authors) that there is any worthwhile intuitive interpretation of this result, its use is to provide a necessary and sufficient criterion to aid the classification of which probability functions do and do not satisfy EP with Ax.

The proof of this theorem, as with all subsequent results in this paper, is given in the Appendix.

We now apply this theorem to our two continua:

Corollary 3.

- *Members of Carnap's Continuum of Inductive Methods,* c_λ^L*, satisfy EP for* $0 < \lambda < \infty$*, and fail to satisfy EP at the endpoints* $\lambda \in \{0, \infty\}$.
- *Members of the Nix-Paris continuum,* w_L^δ*, fail to satisfy EP for* $0 \le \delta \le 1$.

Of course it is well known that c_∞^L fails totally to learn from experience, so its not satisfying EP is hardly a surprise. That c_0^L fails EP is rather for the opposite reason, that it is too ready to jump to the conclusion that all the individuals will be the same as the first one observed. In consequence the corresponding $w_{*\Gamma}$ 'keeps no record' of the g_i, the numbers of each atom instantiated by individuals so far observed, it has no need to because all possible observations are already determined, seen or not seen.

The failure of EP for the w_L^δ for $0 \le \delta \le 1$ is really no surprise given that these probability functions possess the property of *Recovery* (indeed it characterizes them, see [20], [18, Chapter 19]) whereby new observations can effectively erase previous observations. Nevertheless it is strange that this must follow from their characterizing property of satisfying (essentially) Generalized Instantial Relevance, see [13], though the link seems currently explicable only via the underlying mathematics.

The Perspective Principle

In this section we define the Perspective Principle, in contrast to the Elephant Principle a principle of forgetting, and show that the w_L^δ fail to satisfy it for $0 \le \delta \le 1$. We show that satisfying Reichenbach's Axiom is a sufficient condition for a probability function to satisfy the Perspective Principle, from which it follows that the c_λ^L satisfy it for $0 < \lambda < \infty$. The converse fails, a counter-example is c_∞^L.

As remarked in the introduction, the Perspective Principle was originally conceived as a counter-balance to EP, to ensure that where different observations lead to different probability assignments these differences are somehow 'proportional'. The principle is defined, again to avoid any arbitrary measure of similarity, as follows.

The Perspective Principle, PP

Given $\varepsilon > 0$ *and state descriptions* $\Theta(a_1, \dots, a_n), \Phi(a_1, \dots, a_n), \Psi(a_1, \dots, a_r)$*, there is an* m *such that for all state descriptions* $\Xi(a_{n+1}, \dots, a_k)$ *with* $k \ge n + m$,

$$|w(\Psi(a_{k+1}, \dots, a_{k+r}) \mid \Xi \wedge \Theta) - w(\Psi(a_{k+1}, \dots, a_{k+r}) \mid \Xi \wedge \Phi)| < \varepsilon. \qquad (7)$$

Note that in this principle there is the implicit assumption that the probability function w is *Regular*, i.e. $w(\theta) > 0$ for all consistent $\theta \in QFSL$[9], in order that the inequality in (7) makes sense. The c_λ^L and the w_L^δ satisfy Regularity except at the common end point c_0^L, w_L^1 and for this reason this particular member of the continua will not be considered in the remainder of this section.

In essence then, the Perspective Principle says that no matter what observations Θ, Φ we start with, subsequently receiving a sufficiently long stream of common observations Ξ will almost eradicate the significance of this initial difference, at least as far as the probability given to any particular state description Ψ involving just unseen individuals is concerned.

The argument for the rationality of this principle is based on the idea that predictions about future events should be continuous functions of past observations; in that agents who initially adopt the same probability function on the basis of no information should continue to assign similar probabilities if their subsequent observations are sufficiently similar. Put another way it would seem unduly risky (and hence arguably irrational) to adopt a probability function on the basis of no knowledge which could subsequently be critically dependent for all time on the particular properties of a relatively small number of previously observed individuals.

Despite its conception as the twin of EP, PP may be considered a desirable property for the reasons given above, regardless of whether or not the function satisfies EP. We therefore consider the question of whether the w_L^δ satisfy PP, and find on the contrary that

Proposition 4. *For* $0 < \delta < 1$, w_L^δ *fails to satisfy PP.*

In order to present the corresponding classification for the c_λ^L, we need to refer to Reichenbach's Axiom[10].

Reichenbach's Axiom, RA

For w a probability function satisfying Reg,

$$\lim_{m\to\infty}\left(w\left(\alpha_j \mid \bigwedge_{i=1}^{m}\alpha_{h_i}\right) - \frac{u_j\left(\bigwedge_{i=1}^{m}\alpha_{h_i}\right)}{m}\right) = 0 \tag{8}$$

where $u_j\left(\bigwedge_{i=1}^{m}\alpha_{h_i}\right) = |\{i \mid h_i = j\}|$.

[9] Regularity has long been considered to be a rationally desirable property though its strengthening to *Super Regularity*, where we only require $\theta \in SL$, is considerably more contentious, see for example [18, Chapter 10].

[10] See [1] or [18, Chapter 15].

In the appendix we give a proof of the following theorem:

Theorem 5. *If w is a probability function satisfying Reg and RA then w satisfies PP.*

It follows from (3) and (8) that for $0 < \lambda < \infty$ the c_λ^L satisfy Reg and RA, which leads to the result (with $\lambda = \infty$ a trivial case) that

Corollary 6. *For $0 < \lambda \leq \infty$, c_λ^L satisfies PP.*

Conclusion

In this paper we have proposed arguably rational principles of probability assignment based on considerations of never forgetting past observations viz-a-viz predictions about future observations (EP) but at the same time not allowing any such observation to crucially affect these future predictions (PP). These are rather different in nature from the stock symmetry, relevance, irrelevance and analogy considerations which form the basis of most current rational principles in Pure Inductive Logic. Whether they can, or should, have the same force as these stock notions remains open to debate.

We have shown that, for $0 < \lambda < \infty$ the members c_λ^L of Carnap's Continuum satisfy both principles, while all members w_L^δ of the Nix-Paris Continuum for $0 < \delta \leq 1$ (which includes $c_0^L = w_L^1$), satisfy neither, and $c_\infty^L = w_L^0$ satisfies PP without EP. Therefore, if one agrees that these principles are desirable in probability functions used to model rational belief, these results provide support for the choice of the non-extreme (i.e. $\lambda \neq 0, \infty$) c_λ^L over the possible alternatives w_L^δ for such a model.

In this paper we have focused on EP and PP entirely within the classical context of *Unary* Inductive Logic. However these principles would seem to be just as rational (or not) within the more recently developed and wider ambit of *Polyadic* Inductive Logic, where we allow the language to also contain binary, ternary etc. relation symbols. In this case we do have a generalization of de Finetti's Theorem, see [12], [18, Chapter 25] (de Finetti's Theorem is crucial for the proofs in this paper), could this be applied in characterizing EP and PP in non-unary languages?

In addition there remain questions to be answered concerning natural generalizations of EP and PP to all sentences rather than just state descriptions. For example does Theorem 5 continue to hold if we replace the state descriptions $\Theta(\vec{a}), \Phi(\vec{a})$ by just sentences $\theta(\vec{a}), \phi(\vec{a})$ (and use the convention detailed at (6))? For now, these questions remain in the 'in tray'.

References

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

[2] Carnap, R. (1971), A Basic System of Inductive Logic, in *Studies in Inductive Logic and Probability, Volume I*, eds. R. Carnap & R.C. Jeffrey, University of California Press, pp. 33-165.

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

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

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

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

[7] Gaifman, H. (1971), 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, pp. 235-251.

[8] Good, I.J. (1967), On the Principle of Total Evidence, *The British Journal for the Philosophy of Science*, **17**(4): 319-321.

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

[10] Johnson, W.E. (1932), Probability: The Deductive and Inductive Problems, *Mind*, **41**: 409-423.

[11] Kemeny, J.G. (1963), Carnap's Theory of Probability and Induction, in *The Philosophy of Rudolf Carnap*, ed. P.A. Schilpp, La Salle, Illinois, Open Court, pp. 711-738.

[12] Krauss, P.H. (1969), Representation of Symmetric Probability Models, *Journal of Symbolic Logic*, **34**(2): 183-193.

[13] Nix, C.J. & Paris, J.B. (2006), A Continuum of Inductive Methods arising from a Generalized Principle of Instantial Relevance, *Journal of Philosophical Logic*, **35**(1): 83-115.

[14] Paris, J.B., A note on *-conditioning, available at http://www.maths.manchester.ac.uk/~jeff/

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

[16] Paris, J.B. (2011), Pure Inductive Logic, in *The Continuum Companion to Philosophical Logic*, eds. L. Horsten & R. Pettigrew, Continuum International Publishing Group, London, pp. 428-449.

[17] Paris, J.B. & Vencovská, A., Reichenbach's Axiom – For the record. Available at http://www.maths.manchester.ac.uk/~jeff/papers/jp140708.pdf

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

[19] Paris, J.B. & Vencovská, A., The Twin Continua of Inductive Methods, to appear in *Logic Without Borders*, from the Conference in honor of Jouko Väänänen's Sixtieth Birthday, Helsinki, September 2010, Eds. R. Kossak, J. Kennedy, A. Hirvonen & A. Villaveces, published by Ontos Verlag, Frankfurt-Heusenstamm, Germany.

[20] Paris, J.B. & Waterhouse, P. (2009), Atom Exchangeability and Instantial Relevance, *Journal of Philosophical Logic*, **38**(3): 313-332.

Appendix

We prove Theorem 2 via a sequence of lemmas and intermediate discussion.

Recall that $\mathbb{N}_n$ denotes $\{1, 2, \ldots, n\}$ and for $S \subset \mathbb{N}_{2^q}$,

$$N_S = \{\vec{x} \in \mathbb{D}_{2^q} \mid x_i = 0 \Leftrightarrow i \in S\}.$$

Note that these N_S partition $\mathbb{D}_{2^q}$. We shall use S' to denote $\mathbb{N}_{2^q} \setminus S$. For $\vec{x} \in \mathbb{D}_{2^q}$, let $S_{\vec{x}}$ denote the unique S such that $\vec{x} \in N_S$, so $S_{\vec{x}} = \left\{ i \in \mathbb{N}_{2^q} \mid x_i = 0 \right\}$.

Let w be a probability function on SL with de Finetti prior μ, so $w = \int_{\mathbb{D}_{2^q}} w_{\vec{x}} \, d\mu$.

If w does not satisfy EP, there must exist $\Gamma(a_1, \ldots, a_g) = \bigwedge_{i=1}^{2^q} \alpha_i^{g_i}$ and $\Gamma'(a_1, \ldots, a_h) = \bigwedge_{i=1}^{2^q} \alpha_i^{h_i}$ such that $w(\Gamma), w(\Gamma') > 0$ (otherwise EP holds trivially by our convention (6)) and by de Finetti's Theorem 1

$$\frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + n_i} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i} d\mu} = \frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i + n_i} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i} d\mu} \tag{9}$$

for any $n_1, n_2, \ldots, n_{2^q} \in \mathbb{N}$.

Let $M \subseteq \mathbb{D}_{2^q}$ be the set of *support points* of μ. In other words

$$M = \{\vec{x} \in \mathbb{D}_{2^q} \mid \mu(B_\varepsilon(\vec{x})) > 0 \text{ for all } \varepsilon > 0\},$$

where B_ε is the ε neighbourhood $\{\vec{y} \in \mathbb{D}_{2^q} \mid |\vec{x} - \vec{y}| < \varepsilon\}$ of $\vec{x}$. Let G, H be the sets of indices of atoms mentioned in Γ, Γ', respectively, so $G = \{i \in \mathbb{N}_{2^q} \mid g_i > 0\}$ and $H = \{i \in \mathbb{N}_{2^q} \mid h_i > 0\}$ and let G', H' be the complement in $\mathbb{N}_{2^q}$ of G, H respectively, so $G' = \{i \in \mathbb{N}_{2^q} \mid g_i = 0\}$ etc..

Lemma 7. *If w fails EP with* $\Gamma = \bigwedge_{i \in G} \alpha_i^{g_i}$, $\Gamma' = \bigwedge_{i \in H} \alpha_i^{h_i}$ *then*

$$\mu\left(\bigcup_{S \subseteq G' \cap H'} N_S \right) > 0.$$

Proof. Suppose, on the contrary, that w fails EP with Γ, Γ' as described and $\mu(N_S) = 0$ for all $S \subseteq G' \cap H'$. Then by (9), since $\prod_{i \in S} 0^{g_i + n_i} = 0$ whenever $S \cap G \neq \emptyset$ and $\prod_{i \in S} 0^{g_i + n_i} = 1$ whenever $g_i = n_i = 0$,

$$\begin{aligned}
\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + n_i} d\mu &= \sum_{S \subset \mathbb{N}_{2^q}} \int_{N_S} \prod_{i \in S \cap G} 0^{g_i + n_i} \prod_{i \in S \cap G'} 0^{0 + n_i} \prod_{i \in S' \cap G} x_i^{g_i + n_i} \prod_{i \in S' \cap G'} x_i^{0 + n_i} \, d\mu \\
&= \sum_{S \cap G = \emptyset} \int_{N_S} \prod_{i \in S \cap G'} 0^{0 + n_i} \prod_{i \notin S} x_i^{g_i + n_i} \, d\mu \\
&= \sum_{\substack{S \cap G = \emptyset \\ S \cap H \neq \emptyset}} \int_{N_S} \prod_{i \in S \cap G'} 0^{0 + n_i} \prod_{i \notin S} x_i^{g_i + n_i} \, d\mu,
\end{aligned}$$

by our assumption that $\mu(N_S) = 0$ for all $S \subseteq G' \cap H'$. Using the corresponding result for H we obtain from (9) that

$$\begin{aligned}
& w(\Gamma') \left(\sum_{\substack{S \cap G = \emptyset \\ S \cap H \neq \emptyset}} \int_{N_S} \prod_{i \in S \cap G'} 0^{0 + n_i} \prod_{i \notin S} x_i^{g_i + n_i} d\mu \right) \\
&\qquad = w(\Gamma) \left(\sum_{\substack{S \cap G \neq \emptyset \\ S \cap H = \emptyset}} \int_{N_S} \prod_{i \in S \cap H'} 0^{0 + n_i} \prod_{i \notin S} x_i^{h_i + n_i} d\mu \right). \qquad (10)
\end{aligned}$$

Furthermore, it must be the case that

$$\mu \left(\bigcup_{\substack{S \cap G \neq \emptyset \\ S \cap H = \emptyset}} N_S \right) > 0,$$

since otherwise

$$\begin{aligned}
w(\Gamma') &= \sum_{S \subset \mathbb{N}_{2^q}} \int_{N_S} \prod_{i \in S \cap H} 0^{h_i} \prod_{i \in S \cap H'} 0^0 \prod_{i \in S' \cap H} x_i^{h_i} \, d\mu \\
&= \sum_{S \cap H = \emptyset} \int_{N_S} \prod_{i \in S' \cap H} x_i^{h_i} \, d\mu \\
&= \sum_{\substack{S \cap H = \emptyset \\ S \cap G \neq \emptyset}} \int_{N_S} \prod_{i \in S' \cap H} x_i^{h_i} d\mu = 0
\end{aligned}$$

(again by the assumption that $\mu(N_S) = 0$ for all $S \subseteq G' \cap H'$), contradicting $w(\Gamma') > 0$.

Therefore, letting $n_i > 0$ for all $i \in H \cap G'$ and $n_i = 0$ for all $i \in G \cap H'$ gives a value of 0 on the left of (10) with a positive value on the right, contradicting (9). The result follows. □

Lemma 8. *If w fails EP with* $\Gamma = \bigwedge_{i \in G} \alpha_i^{g_i}$, $\Gamma' = \bigwedge_{i \in H} \alpha_i^{h_i}$, *then for any* $\vec{d} \in M$ *such that* $S_{\vec{d}} \subseteq G' \cap H'$, *and any* $\vec{c} \in M$

$$\prod_{i=1}^{2^q} c_i^{g_i} d_i^{h_i} = \prod_{i=1}^{2^q} c_i^{h_i} d_i^{g_i}. \tag{11}$$

Proof. Suppose $\vec{d} = \langle d_1, d_2, \ldots, d_{2^q} \rangle \in M$ is such that $S_{\vec{d}} \subseteq G' \cap H'$ and $\vec{c} = \langle c_1, c_2, \ldots, c_{2^q} \rangle \in M$. Let $n \in \mathbb{N}$ be large, then letting n_i in (9) take values $[nc_i]$, $[nd_i]$ in turn, where $[x]$ denotes the integer part of $x \in \mathbb{R}$, and dividing the first equation obtained by the second obtained gives

$$\frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + [nc_i]} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + [nd_i]} d\mu} = \frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i + [nc_i]} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i + [nd_i]} d\mu}. \tag{12}$$

Since $\vec{d} \in M$, $\mu(B_\varepsilon(\vec{d})) > 0$ for any $\varepsilon > 0$. Let $0 < \varepsilon < min\{d_i \mid d_i > 0\}$ and let $\vec{x} \in B_\varepsilon(\vec{d})$. Then $S_{\vec{x}} \subseteq S_{\vec{d}}$, for otherwise there must exist some i such that $x_i = 0 < d_i$, giving $|\vec{x} - \vec{d}| \geq \sqrt{d_i^2} > \varepsilon$, a contradiction. Therefore, if $\vec{d} \in M$ then $\mu\left(\bigcup_{S \subseteq S_{\vec{d}}} N_S\right) > 0$, so for $T = \bigcup_{S \subseteq S_{\vec{d}}} N_S$,

$$\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + [nd_i]} d\mu \geq \int_T \prod_{i \in S_{\vec{d}}} x_i^{g_i + 0} \prod_{i \notin S_{\vec{d}}} x_i^{g_i + [nd_i]} \, d\mu > 0$$

since $d_i = 0$ for all $i \in S_{\vec{d}}$. Likewise $\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i + [nd_i]} d\mu > 0$, so (12) is well-defined.

Dividing both sides of (12) by

$$\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{[nd_i]} d\mu \cdot \int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{[nc_i]} d\mu,$$

which is similarly well-defined, and rearranging gives

$$\frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + [nc_i]} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{[nc_i]} \, d\mu} \cdot \frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i + [nd_i]} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{[nd_i]} d\mu} = \frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i + [nd_i]} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{[nd_i]} \, d\mu} \cdot \frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i + [nc_i]} d\mu}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{[nc_i]} \, d\mu}.$$

By [9, Lemma 7], taking the limit as $n \to \infty$ then gives

$$\prod_{i=1}^{2^q} c_i^{g_i} d_i^{h_i} = \prod_{i=1}^{2^q} c_i^{h_i} d_i^{g_i}.$$

□

Furthermore, whenever $S_{\vec{c}} \subseteq G' \cap H'$, both sides of (11) are positive, and it is equivalent to

$$\prod_{i=1}^{2^q} c_i^{k_i} = \prod_{i=1}^{2^q} d_i^{k_i} \tag{13}$$

where $k_i = g_i - h_i$. (If $S_{\vec{c}} \not\subseteq G' \cap H'$ then both sides of (11) are zero).

Let S_n denote the set of permutations of $\mathbb{N}_n$.

Lemma 9. *If w satisfies Ax and fails EP with* $\Gamma = \bigwedge_{i \in G} \alpha_i^{g_i}$, $\Gamma' = \bigwedge_{i \in H} \alpha_i^{h_i}$, *then for any* $S \subset \mathbb{N}_{2^q}$ *such that* $|S| \leq |G' \cap H'|$, *there is some constant* X_S *such that*

$$\mu\{\vec{x} \in N_S \mid \prod_{i \notin S} x_i = X_S\} = \mu(N_S). \tag{14}$$

Proof. Let w, S be as described and assume that $\mu(N_S) > 0$, since otherwise (14) holds trivially. Note that for w satisfying Ax μ will be invariant under permutations of the 2^q coordinates (see [19, Chapter 14]), so that for τ a permutation of $\{1, \ldots, 2^q\}$ and A a Borel subset of $\mathbb{D}_{2^q}$

$$\mu(A) = \mu(\tau(A)) = \mu(\{\tau(\vec{x}) \mid \vec{x} \in A\})$$

where $\tau(\vec{x}) = \left\langle x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(2^q)} \right\rangle$.

If $|S| = 2^q - 1$, then N_S is a singleton and the result follows. Otherwise by Ax, since μ is invariant under permutations of co-ordinates, as $|S| \leq |G' \cap H'|$, there must exist $T \subseteq G' \cap H'$ with $|T| = |S|$ and $\mu(N_T) = \mu(N_S) > 0$. Let $\vec{d} \in M \cap N_T$. Let $r, s \in T'$ with $r \neq s$ and let $\sigma \in \mathrm{S}_{2^q}$ be the permutation which exchanges r and s and leaves all other values unchanged. Then $\sigma(\vec{d}) = \langle d_{\sigma(1)}, \ldots, d_{\sigma(2^q)} \rangle$ is also in $M \cap N_T$ by the symmetry of μ. Since w does not satisfy EP then by (13),

$$d_r^{k_r} d_s^{k_s} \prod_{i \neq r, s} d_i^{k_i} = d_s^{k_r} d_r^{k_s} \prod_{i \neq r, s} d_i^{k_i}$$

and therefore

$$\left(\frac{d_r}{d_s}\right)^{k_r} = \left(\frac{d_r}{d_s}\right)^{k_s}$$

giving either $d_r = d_s$ or $k_r = k_s$. For each pair of co-ordinates in T', the permutation exchanging these while leaving all others unchanged may be used similarly to show that, for all $r, s \in T'$, either $d_r = d_s$ (so $d_i = 0$ for $i \in T$ and $d_i = |T'|^{-1}$ for $i \in T'$ is the sole support point of μ in N_T) or $k_r = k_s$ and hence for all $\vec{c}, \vec{d} \in M \cap N_T$

$$\prod_{i \notin T} c_i = \prod_{i \notin T} d_i.$$

In either case, (14) holds for N_T. Let $\tau \in S_{2^q}$ be a permutation such that $\tau(i) \in T \Leftrightarrow i \in S$. Then by Ax

$$\vec{x} \in M \cap N_S \Rightarrow \tau(\vec{x}) \in M \cap N_T \Rightarrow \prod_{i \notin T} x_{\tau(i)} = X_T = \prod_{i \notin S} x_i,$$

and so, again by the symmetry of μ and since $\mu(M \cap N_S) = \mu(N_S)$,

$$\mu(\{\vec{x} \in N_S \mid \prod_{i \notin S} x_i = X_T\}) = \mu(\{\tau(\vec{y}) \mid \vec{y} \in N_T, \prod_{i \notin T} y_i = X_T\}) = \mu(N_T) = \mu(N_S).$$

□

We are now in a position to prove:

Theorem 2. *Suppose that w is a probability function satisfying Ax with de Finetti prior μ, and let $z = \min\{|S|$ such that $\mu(N_S) > 0\}$. Then w fails EP just if there is some $X \in \mathbb{R}$ such that*

$$\mu(\{\vec{x} \in N_S \mid \prod_{i \notin S} x_i = X\}) = \mu(N_S),$$

for every $S \subset \mathbb{N}_{2^q}$ such that $|S| = z$.

Proof. Suppose w, μ and z are as described. Suppose firstly that there is some $X \in \mathbb{R}$ such that $\mu(\{\vec{x} \in N_S \mid \prod_{i=1}^{2^q} x_i = X\}) = \mu(N_S)$ for every $S \subset \mathbb{N}_{2^q}$ such that $|S| = z$. Let $T \subset \mathbb{N}_{2^q}$ with $|T| = z$, so that $\mu(N_T) > 0$ while $\mu(N_S) = 0$ whenever $|S| < |T|$, and for all $\vec{d} \in M \cap N_T$, $\prod_{i \notin T} d_i = X$. Let $\Gamma = \bigwedge_{i \notin T} \alpha_i^g$, $\Gamma' = \bigwedge_{i \notin T} \alpha_i^h$ for some $g, h \in \mathbb{N}$ with $g, h, > 0$, $g \neq h$. Then

$$\frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i+n_i}\, d\mu(\vec{x})}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{g_i}\, d\mu(\vec{x})} = \frac{\int_{N_T} \prod_{i\in T} 0^{0+n_i} \prod_{i\not\in T} x_i^{g+n_i}\, d\mu(\vec{x})}{\int_{N_T} \prod_{i\in T} 0^{0} \prod_{i\not\in T} x_i^{g}\, d\mu(\vec{x})}$$

$$= \frac{X^g \int_{N_T} \prod_{i=1}^{2^q} x_i^{n_i}\, d\mu(\vec{x})}{X^g \int_{N_T} 1\, d\mu(\vec{x})}$$

$$= \frac{1}{\mu(N_T)} \int_{N_T} \prod_{i=1}^{2^q} x_i^{n_i}\, d\mu(\vec{x})$$

since for every $S \subset T$, $\mu(N_S) = 0$ and for every $S \not\subseteq T$, each $\vec{x} \in N_S$ has some zero co-ordinate $x_i = 0$ with $i \not\in T$, so that $\prod_{i\not\in T} x_i^g = 0$. Substituting h for g shows that

$$\frac{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i+n_i}\, d\mu(\vec{x})}{\int_{\mathbb{D}_{2^q}} \prod_{i=1}^{2^q} x_i^{h_i}\, d\mu(\vec{x})}$$

takes the same value, so w fails EP.

In the other direction, suppose w fails EP with $\Gamma = \bigwedge_{i\in G} \alpha_i^{g_i}$, $\Gamma' = \bigwedge_{i\in H} \alpha_i^{h_i}$. By Ax (and the associated symmetry of μ), $\mu(N_S) > 0$ for every $S \subset \mathbb{N}_{2^q}$ of size z, and by Lemma 7, $z \le |G' \cap H'|$, so there is some such $S \subseteq G' \cap H'$. Therefore, by Lemma 9, the result follows. □

This result gives a complete characterization of the probability functions satisfying Ax with EP, from which we obtain the following

Corollary 3.

- *Members of Carnap's Continuum, c_λ^L, satisfy EP for $0 < \lambda < \infty$, and fail to satisfy EP at the endpoints $\lambda \in \{0, \infty\}$.*
- *Members of the Nix-Paris continuum w_L^δ fail to satisfy EP for $0 \le \delta \le 1$.*

Proof. That the c_λ^L satisfy EP for $0 < \lambda < \infty$ follows from Theorem 2 using the fact that, from (4), every point in $\mathbb{D}_{2^q}$ is a support point of c_λ^L for $0 < \lambda < \infty$. So for $\vec{x} = \langle 2^{-q}, 2^{-q}, \ldots, 2^{-q} \rangle$, $\vec{y} \in N_\emptyset$ with $\vec{y} \ne \vec{x}$, both $\vec{x}, \vec{y} \in M \cap N_\emptyset$ but their co-ordinate products are not equal, since $\prod_{i=1}^{2^q} z_i$ has a strict maximum at $\vec{x}$ for $\vec{z} \in \mathbb{D}_{2^q}$.

That EP fails to hold for the w_L^δ for $0 \le \delta \le 1$ (which includes $c_0^L = w_L^1$ and $c_\infty^L = w_L^0$) follows from Theorem 2 using the observation that by (5), in each case, the support points of the de Finetti prior are all permutations of each other:

$$M = \{\vec{e}_1(\delta), \vec{e}_2(\delta), \ldots, \vec{e}_{2^q}(\delta)\} = \{\sigma(\vec{e}_1(\delta)) \mid \sigma \in S_{2^q}\},$$

so all have the same co-ordinate product. □

Proposition 4. *For* $0 < \delta < 1$*,* w_L^δ *fails to satisfy PP.*

Proof. From the definitions given above in (5) and (1), it follows that

$$w_L^\delta\left(\bigwedge_{i=1}^m \alpha_{h_i}\right) = 2^{-q}\sum_{j=1}^{2^q} \gamma^{m-m_j}(\gamma+\delta)^{m_j} \tag{15}$$

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

Let $n \in \mathbb{N}^+$ and let $\Theta(a_1,\ldots,a_n) = \bigwedge_{i=1}^{2^q} \alpha_i^{t_i}$, and $\Phi(a_1,\ldots,a_n) = \bigwedge_{i=1}^{2^q} \alpha_i^{p_i}$ be state descriptions for $a_1,\ldots,a_n$. For any $m \in \mathbb{N}$, choose $h \ge m$ such that $h = g2^q$ for some $g \in \mathbb{N}$. Let $\Xi(a_{n+1},\ldots,a_{n+h}) = \bigwedge_{i=1}^{2^q} \alpha_i^g$. Let $r \in \mathbb{N}^+$ and let $\Psi(a_{n+h+1},\ldots,a_{n+h+r}) = \bigwedge_{i=1}^{2^q} \alpha_i^{r_i}$ be a state description for $a_{n+h+1},\ldots,a_{n+h+r}$. Then by (15)

$$
\begin{aligned}
&|w_L^\delta(\Psi \mid \Xi \wedge \Theta) - w_L^\delta(\Psi \mid \Xi \wedge \Phi)| \\
&= \left| \frac{\sum_{i=1}^{2^q} \gamma^{r+h+n-(r_i+g+t_i)}(\gamma+\delta)^{r_i+g+t_i}}{\sum_{i=1}^{2^q} \gamma^{h+n-(g+t_i)}(\gamma+\delta)^{g+t_i}} - \frac{\sum_{i=1}^{2^q} \gamma^{r+h+n-(r_i+g+p_i)}(\gamma+\delta)^{r_i+g+p_i}}{\sum_{i=1}^{2^q} \gamma^{h+n-(g+p_i)}(\gamma+\delta)^{g+p_i}} \right| \\
&= \left| \frac{\sum_{i=1}^{2^q} \gamma^{r+n-(r_i+t_i)}(\gamma+\delta)^{r_i+t_i}}{\sum_{i=1}^{2^q} \gamma^{n-t_i}(\gamma+\delta)^{t_i}} - \frac{\sum_{i=1}^{2^q} \gamma^{r+n-(r_i+p_i)}(\gamma+\delta)^{r_i+p_i}}{\sum_{i=1}^{2^q} \gamma^{n-p_i}(\gamma+\delta)^{p_i}} \right| \\
&= |w_L^\delta(\Psi \mid \Theta) - w_L^\delta(\Psi \mid \Phi)|.
\end{aligned}
$$

For $\delta > 0$, Θ, Φ and Ψ may be chosen such that this last value is greater than 0. Therefore, the value of $|w_L^\delta(\Psi \mid \Xi \wedge \Theta) - w_L^\delta(\Psi \mid \Xi \wedge \Phi)|$ is fixed, positive and independent of the value of h, which may be arbitrarily large, and w_L^δ fails PP. (For $\delta = 0$ the value will always be zero so $w_L^0 (= c_\infty^L)$ trivially satisfies PP.) □

Theorem 5. *If w is a regular probability function satisfying RA then w satisfies PP.*

Since both RA and PP seem to capture in some sense the idea of the probability function following, and ultimately converging to, the objective frequencies in the observations, one might have hoped that Theorem 5 would have a fairly elementary and transparent proof. However we currently know of no such proof, the one we present depends on the following technical characterization of the probability functions satisfying RA due to Haim Gaifman [7].

Theorem 10. *Let w satisfy Reg. Then w satisfies RA if and only if every point in $\mathbb{D}_{2^q}$ is a support point of the de Finetti prior of w.*

Proof of Theorem 5. It follows from the proof of Theorem 10 given in [18, Chapter 15] (or see [17, Corollory 2]) that if RA holds then it holds uniformly, so that for any $\nu > 0$ there is some $t \in \mathbb{N}$ such that for any sequence of atoms α_{g_i} for $i = 1, \ldots, m$ with $m \geq t$,

$$\left| w\left(\alpha_j \mid \bigwedge_{i=1}^{m} \alpha_{g_i}\right) - \frac{u_j\left(\bigwedge_{i=1}^{m} \alpha_{g_i}\right)}{m} \right| < \nu. \tag{16}$$

We shall need this ‘stronger version’ of RA in what follows.

Suppose w is a regular probability function which satisfies RA. Let $n \in \mathbb{N}^+$ and let $\Theta(a_1, \ldots, a_n), \Phi(a_1, \ldots, a_n)$ be arbitrary fixed state descriptions for $a_1, \ldots, a_n$. Let $r \in \mathbb{N}^+$ and $\Psi(a_1, \ldots, a_r) = \bigwedge_{i=1}^{r} \alpha_{s_i}$ be an arbitrary fixed state description for $a_1, \ldots, a_r$. Let $m \in \mathbb{N}$ and let $k = n + m$. Then for any $\Xi(a_{n+1}, \ldots, a_k)$

$$\begin{aligned}
&\left| w(\Psi(a_{k+1}, \ldots, a_{k+r}) \mid \Xi \wedge \Theta) - w(\Psi(a_{k+1}, \ldots, a_{k+r}) \mid \Xi) \right| \\
&= \left| w\left(\bigwedge_{i=1}^{r} \alpha_{s_i} \mid \Xi \wedge \Theta\right) - w\left(\bigwedge_{i=1}^{r} \alpha_{s_i} \mid \Xi\right) \right| \\
&= \left| \prod_{b=1}^{r} w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right) - \prod_{b=1}^{r} w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right) \right|.
\end{aligned} \tag{17}$$

For any fixed $b \in \{1,\ldots,r\}$,

$$\left| w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right) - w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right)\right|$$

$$\leq \left| w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right) - \frac{u_{s_b}\left(\bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right)}{k+b-1}\right|$$

$$+ \left| w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right) - \frac{u_{s_b}\left(\bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right)}{m+b-1}\right|$$

$$+ \left| \frac{u_{s_b}\left(\bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right)}{k+b-1} - \frac{u_{s_b}\left(\bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right)}{m+b-1}\right|. \tag{18}$$

By (16) and since

$$u_{s_b}\left(\bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right) = u_{s_b}\left(\bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right) + u_{s_b}(\Theta)$$

where $u_{s_b}(\Theta) \leq n$, (18) is smaller than any given $\delta > 0$, provided that m is taken sufficiently large.

Let $P_b = \min\{w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi\right), w\left(\alpha_{s_b} \mid \bigwedge_{i=1}^{b-1} \alpha_{s_i} \wedge \Xi \wedge \Theta\right)\} \leq 1$, then if δ is an upper bound for (18), the value of (17) is less than

$$\prod_{b=1}^{r}(P_b + \delta) - \prod_{b=1}^{r} P_b \leq \delta r + \delta^2 \binom{r}{2} + \cdots + \delta^r.$$

Given any $\varepsilon > 0$, δ may be chosen such that the above is less than $\varepsilon/2$. The same argument may also be used with Φ in place of Θ to finally obtain

$$\Big| w(\Psi(a_{k+1},\ldots,a_{k+r}) \mid \Xi(a_{n+1},\ldots,a_k) \wedge \Theta(a_1,\ldots,a_n))$$
$$- w(\Psi(a_{k+1},\ldots,a_{k+r}) \mid \Xi(a_{n+1},\ldots,a_k) \wedge \Phi(a_1,\ldots,a_n)) \Big| < \varepsilon$$

for any $\Xi(a_{n+1},\ldots,a_k)$ where k is sufficiently large. Therefore, w satisfies PP. □

Elizabeth HOWARTH and Jeff B. PARIS
School of Mathematics
The University of Manchester
Manchester M13 9PL
lizhowarth@outlook.com
jeff.paris@manchester.ac.uk