Analogical Predictive Probabilities

Simon M. Huttegger
University of California, Irvine
shuttegg@uci.edu

It is well known that Rudolf Carnap's original system of inductive logic failed to provide an adequate account of analogical reasoning. Since this problem was identified, there has been no shortage of proposals for how to incorporate analogy into inductive inference. Most alternatives to Carnap's system, unlike his original one, have not been derived from first principles; this makes it to some extent unclear what the epistemic situations are to which they apply. This paper derives a new analogical inductive logic from a set of axioms which extend Carnap's postulates in a natural way. The key insights come from Bruno de Finetti's ideas about analogy. The axioms of the new system capture epistemic conditions that call for a strong kind of analogical reasoning. The new system has a number of merits, but is also subject to limitations. I shall discuss both, together with some possible ways to generalize the approach taken in this paper.

The Bayesian version of the simplest form of inductive inference—induction by enumeration—equates the predictive probability of an outcome with its observed relative frequency (up to initial beliefs). One approach for uncovering the principles that underlie this method was pioneered by the Cambridge logician and philosopher W. E. Johnson (1924, 1932). It was later developed independently by Rudolf Carnap (1950, 1952, 1971, 1980) in his epic project of reconstructing the inductive logic of science. At the heart of the Johnson-Carnap approach is an application of the axiomatic method: their inductive systems are derived from a set of rather plausible axioms, which require in effect that the order in which outcomes are observed is probabilistically irrelevant. The Johnson-Carnap approach thus shows that induction by enumeration is adequate whenever the order of events is judged to be of no significance.[1]

A limitation of the Johnson-Carnap approach is its inability to deal with various forms of analogical reasoning. This was pointed out by Peter Achinstein (1963) in an influential critique of Carnap's inductive system. Achinstein showed that Carnapian inductive logic cannot

[1] See Zabell (2011) for an overview. For more information about the role of order invariance see Zabell (1989).

Mind, Vol. 128 . 509 . January 2019
doi:10.1093/mind/fzw049 Advance Access publication 8 June 2017

account for the following plausible inference pattern: if all types of metal you have observed so far conduct electricity, a hitherto unobserved type of metal will very likely exhibit the same behaviour. Carnap (1963) clearly recognized the problem. But while his original system of inductive logic includes some limited forms of analogical inference (Carnap and Stegmüller 1959; Carnap 1980; Niiniluoto 1981; Romeijn 2006), Carnap's own contributions did not lead to any definitive solution. The reason is that the most interesting kinds of analogy are excluded by one of the central principles of Johnson's and Carnap's inductive logic, the so-called sufficientness postulate (Good 1965). The sufficientness postulate implies that the predictive probability of an outcome is independent of how often similar but distinct outcomes have been observed. Carnapian inductive logic is hostile ground for analogy at its most fundamental level.

There have been a number of attempts to solve this problem.[2] In this paper I develop a novel inductive system that departs from previous accounts in a number of significant ways. The new inductive logic extends Carnap's system so as to enable distinct but similar *types* of observations to influence each other's predictive probabilities. It is based on two generalizations of Johnson's and Carnap's axioms. One lifts the restrictions imposed by the sufficientness postulate. The other one is a weakening of order invariance along the lines suggested by Bruno de Finetti. De Finetti's early work on inductive inference also used order invariance, which in modern terminology is known as *exchangeability*. Exchangeability, according to de Finetti, is a judgement to the effect that the trials of an experiment are fully analogous. In order to allow for analogies that are less than complete, de Finetti (1938, 1959) introduced the concept of *partial exchangeability*, which requires invariance only under restricted classes of reorderings. Besides a generalized sufficientness postulate, a new variant of partial exchangeability will be the most important element of the new inductive system.

After a brief review of exchangeability, partial exchangeability is introduced in §1. In §2 I state the main result and discuss some of its consequences. Finally, in §3 I analyse the asymptotic behaviour of the new inductive system and examine it in the context of a famous principle, known as 'Reichenbach's axiom', which requires predictive

[2] See Hesse (1964), Niiniluoto (1981), Spohn (1981), Costantini (1983), Kuipers (1984), Skyrms (1993a), di Maio (1995), Festa (1997), Maher (2000, 2001), Romeijn (2006), and Hill and Paris (2013).

probabilities to converge to limiting relative frequencies. A discussion of the issues revolving around Reichenbach's axiom leads to an increased understanding of the conditions under which the inductive logic of this paper is an adequate model.

1. Exchangeability and partial exchangeability

1.1 Exchangeability

The concept of exchangeability is at the heart of the Bayesian approach to inductive inference; it provides us with a Bayesian version of the most common type of statistical model. A sequence of outcomes is exchangeable if its probability does not depend on the order in which outcomes are observed; reordering the sequence has no effect on its probability. Mathematically, a finite sequence of outcomes is represented by a finite sequence of random variables, $X_1, \ldots, X_n$; for simplicity we assume that each X_i takes values in a finite set of outcomes, $\{1, \ldots, s\}$. The sequence $X_1, \ldots, X_n$ is *exchangeable* if

$$\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \mathbb{P}[X_1 = x_{\sigma(1)}, \ldots, X_n = x_{\sigma(n)}]$$

for every permutation σ of $\{1, \ldots, n\}$.[3] Exchangeability can be extended to infinitely many observations: an infinite sequence $X_1, X_2, \ldots$ of outcomes is exchangeable whenever each of its finite initial segments is exchangeable.

That infinite exchangeable sequences fully capture the most important case of statistical inference is a consequence of de Finetti's famous representation theorem.[4] De Finetti showed that infinite exchangeable sequences are equivalent to independent and identically distributed (i.i.d.) trials of a chance experiment with unknown chances. Thus, whenever $X_1, X_2, \ldots$ is an infinite exchangeable sequence, its associated probability measure $\mathbb{P}$ can be represented by a unique mixture of i.i.d. multinomial trials; conversely, every such mixture determines an exchangeable probability over sequences of observations.[5] As a result,

[3] Such a permutation is a reordering of the numbers $1, \ldots, n$.

[4] See de Finetti (1937). There are many extensions of de Finetti's result to more general probability spaces; see Aldous (1985) for an overview.

[5] Suppose, for example, that X_n records whether the nth toss of a coin flip comes up heads or tails, and that the infinite sequence $X_1, X_2, \ldots$ is exchangeable. According to de Finetti's representation theorem this is equivalent to the probability of finite sequences of heads and

observations may be thought of as generated by a process according to which each outcome has a fixed probability, or chance, of occurring. Each vector $(p_1, \ldots, p_s)$, where p_i is the chance of outcome i and $\sum_{i=1}^{s} p_i = 1$, is a chance configuration. If chances are unknown, uncertainty can be expressed as a probability distribution over the set of all possible chance configurations, $(p_1, \ldots, p_s)$, which is the *mixing measure* in the de Finetti representation. De Finett's theorem thus allows us to view *unobservable* chance set-ups as intrinsic features of the probabilistic structure of the *observable* process. If an agent's subjective degrees of belief are exchangeable, that agent is free to adopt any method that refers to the chance set-up without having to think of chances as real. De Finetti's representation theorem entitles subjective Bayesians to help themselves to chances whenever it's useful to do so.

One important consequence of de Finetti's representation theorem is a simple formula for calculating predictive probabilities. Suppose the mixing measure in the de Finetti representation is a Dirichlet distribution.[6] Then the conditional probability of observing outcome i given past observations is

$$(2) \quad \mathbb{P}[X_{n+1} = i | X_1, \ldots, X_n] = \frac{n_i + \alpha_i}{n + \sum_i \alpha_i}$$

where n_i is the number of times i occurs among $X_1, \ldots, X_n$, and the parameters $\alpha_j, 1 \leq j \leq s,$ determine the initial probablilities of observing outcomes. Formula (2) is a generalization of *Laplace's rule of succession* (Laplace 1774), which sets all alpha parameters equal to one.

The derivation of (2) from de Finetti's theorem appeals to special prior distributions over chance set-ups. There is an alternative approach to generalized rules of succession that does not rely on

tails being a mixture of i.i.d. binomial trials with unknown bias of the coin:

$$(1) \quad \mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \int_0^1 p^h (1-p)^{n-h} d\mu(p)$$

Here, p is the bias for heads, μ is a uniquely determined prior over biases, and h is the number of heads in the first n trials.

[6] This means that the mixing measure is given by

$$\frac{\Gamma\left(\sum_{j=1}^{K} \alpha_j\right)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} \, p_1^{\alpha_1 - 1} \ldots p_K^{\alpha_K - 1} dp_1 \ldots dp_{K-1}$$

where Γ is the gamma function. By varying the alpha parameters, Dirichlet distributions are capable of expressing many different states of uncertainty.

such assumptions. This approach was developed independently by W. E. Johnson (1924, 1932) and Rudolf Carnap (1950, 1952, 1971, 1980). Rather than using chance set-ups, they introduced what has become known as *Johnson's sufficientness postulate*, which requires the predictive probability of i to depend only on i, n_i, and the total sample size n:

(3) $\mathbb{P}[X_{n+1} = i | X_1, \ldots, X_n] = f_i(n_i, n)$

The predictive probabilities (2) can be derived from Johnson's sufficientness postulate if we assume, in addition, that the sequence of observations is exchangeable.[7]

What this tells us is that, in the presence of exchangeability, Johnson's sufficientness postulate characterizes Dirichlet priors. This is very much in line with a subjective Bayesian point of view such as that of de Finetti, who maintains that the chance set-up of the de Finetti representation is just an artefact of exchangeable probabilities; because the chance setup is unobservable, any appeal to it, such as requiring the mixing measure to be Dirichlet, is philosophically problematic.[8] Johnsons's sufficientness postulate gets around this difficulty in a way that only refers to the observable process. Instead of figuring out what you think about chances, all you need to do is judge whether predictive probabilities really depend on nothing but relative frequencies.

1.2 Analogy

Both exchangeability and Johnson's sufficientness postulate have to do with analogy. Let's consider Johnson's sufficientness postulate first. It implies that the number of trials n_k of outcomes k not identical but similar to i is judged to be irrelevant for the conditional probability of i. As observed by Achinstein (1963), this creates a problem for Carnapian inductive logic. Achinstein puts forward an example in which an investigator explores whether different kinds of metal conduct heat. Suppose that so far we have only observed the number of times platinum or osmium conduct heat. On the next trial we are going to observe a piece of rhodium. Should our observations of platinum and osmium be relevant for our predictive probabilities? Analogical reasoning of this type seems to be eminently plausible, because platinum, osmium and rhodium are in the same chemical

[7] See Zabell (1982) and Kuipers (1978) for detailed accounts of the mathematical aspects of, respectively, Johnson's and Carnap's arguments.

[8] See de Finetti (1938).

family. But evidence from observations of platinum and rhodium is exactly what Johnson's sufficientness postulate deems irrelevant. As a result, Carnap's inductive logic cannot deal with these very simple types of analogical inference.

As first observed by de Finetti, exchangeability involves another kind of analogy judgement. De Finetti was fond of noting that each observation or each trial of an experiment is, strictly speaking, unique; the circumstances of distinct observations may vary in any number of ways, and the same kind of observation can be, and usually is, made under completely different background conditions (de Finetti 1938, 1974). As an illustration, consider flipping a coin: the ambient temperature, the time of day, the person who flips the coin, the stock market indices at the time of the coin flip—all these factors vary, perhaps ever so slightly, with each flip of the coin. Judging a sequence of outcomes to be exchangeable is therefore tantamount to asserting that all these background factors are irrelevant; trials are thought to be *completely analogous* in all relevant respects.

Of course, trials often fail to be completely analogous. For example, the coin may be flipped by two people. If one of them is a professional coin-flipper and the other an amateur like me, thinking of trials as exchangeable would not be adequate. De Finetti's standard example is flipping two coins which don't look exactly the same. In order to allow varying circumstances to have an effect on inductive inference, de Finetti (1938, 1959) introduced the concept of *partial exchangeability*. Partial exchangeability captures a middle ground between complete analogy (exchangeability) and no analogy among trials.

1.3 Partial exchangeability

The basic scheme of partial exchangeability involves *outcomes* and *types of outcome*. Suppose there are $t < \infty$ types of outcomes. Let X_{nj} be the nth outcome of type j, N_j the total number of type j outcomes observed thus far, and $N = \sum_j N_j$ the total number of observations. Observations are given by an array that has $X_{1j}, \ldots, X_{N_j,j}$ as its jth column, where $j = 1, \ldots, t$. De Finetti's example of flipping two coins has two types—the coins—and two outcomes—heads and tails. This gives rise to an array of two sequences of heads and tails, one for the first coin and the other for the second coin. Achinstein's example also fits this scheme; different kinds of metal correspond to types, and whether or not an observed metal conducts heat are the outcomes.

A sequence of observations of different types is *partially exchangeable* if it is exchangeable within each type. More precisely, let

$X_{1j}, \ldots, X_{N_j,j}, 1 \leq j \leq t$, be an array of observations, and let n_{ij} be the number of times an outcome i of type j has been observed in the first N trials. Then the array $X_{1j}, \ldots, X_{N_j,j}, 1 \leq j \leq t$, is partially exchangeable if every array with the same counts n_{ij} has the same probability. An infinite array $X_{1j}, X_{2j} \ldots, 1 \leq j \leq t$, is partially exchangeable if every finite initial array is partially exchangeable. This means that reordering finitely many outcomes within a type has no effect on the probability assignment. The numbers n_{ij} are a *sufficient statistic* for the prior probability.

Exchangeability is a special case of partial exchangeability. If probability assignments are not just invariant under reordering outcomes within types (as required by partial exchangeability) but also across types, the probability assignment is exchangeable. This is a precise statement of the idea that exchangeability represents full analogy, whereas partial exchangeability allows weaker analogies between types.

Partially exchangeable probabilities can also be represented in terms of chance set-ups. The appropriate chance set-up is slightly different from the one corresponding to exchangeable probabilities (i.i.d. trials). Trials are independent, but they are identically distributed only within each type; across types, the chances of outcomes can vary. De Finetti's representation theorem for partially exchangeable probabilities says that this chance set-up, together with a prior over chances, is equivalent to partially exchangeable probability measures.[9] In de Finetti's coin example, the representation theorem allows us to think of the partially exchangeable process as generated by first choosing biases p and q of the coins according to a joint distribution over biases and then flipping the two coins independently with probabilities p and q for heads.

Notice that the chance set-up allows p and q to be unequal but correlated. This allows a continuum of analogy effects between types. If biases are perfectly correlated, they are equal and we have exchangeability or full analogy. If biases are chosen independently, there is no analogy between types. In general, though, types may be arbitrarily

[9] See de Finetti (1938, 1959), Link (1980), and Diaconis and Freedman (1980). More formally, if each type occurs infinitely often, then there exists a unique measure μ such that for all $0 \leq n_{ij} \leq N_j, i = 1, \ldots, s, j = 1, \ldots, t$,

$$\begin{aligned} &\mathbb{P}[X_{11} = x_{11}, \ldots, X_{N_1,1} = x_{N_1,1}; \ldots; X_{1t} = x_{1t}, \ldots, X_{N_t,t} = x_{N_t,t}] \\ &\quad = \int_{\Delta^t} \prod_{j=1}^{t} p_{1j}^{n_{1j}} \cdots p_{sj}^{n_{sj}} \, d\mu(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_t) \end{aligned} \tag{4}$$

The integral ranges over the t-fold product of the $s-1$-dimensional unit simplex Δ, and μ is the mixing measure on the probability vectors $\boldsymbol{p}_j = (p_{1j}, \ldots, p_{sj}) \in \Delta, 1 \leq j \leq t$.

correlated. This correlation creates analogy influences between types in de Finetti's framework.[10]

In order to understand better the type of analogy which is at play here, it is instructive to consider the limiting behaviour of partially exchangeable processes. Regardless of the degree of correlation between types, the posterior distribution becomes peaked around the observed relative frequencies with increasing information. This implies that the predictive probabilities, which coincide with the Bayes estimates (the posterior expectation of chances), converge to the limiting relative frequencies with probability one. For example, the predictive probabilities of outcomes when flipping two coins almost certainly converge to the biases p and q. From this we can infer that de Finetti's analogies are transient; similarities between types have an effect only on how types are initially correlated.

I emphasize this point because the inductive system I will introduce in the next section models a type of analogy that is permanent and not transient. Let me give two examples of when we should expect such persistent correlations. The first is a set of clinical trials with male and female patients (types) who can display different symptoms (outcomes) that depend on unknown infectious agents. In this situation, it is reasonable to form different conditional beliefs about observing symptoms in future subjects according to whether they are females or males. If there is an underlying chance process (how infectious agents are expressed) according to which symptoms in female and male subjects covary, observations of one type might not cease to provide information about the other type as the sample size grows without bound. This may result in enduring analogy effects.

The second example comes from game theory. The standard solution concept of game theory, Nash equilibrium, assumes that players choose their strategies independently. In many games this is an implausible assumption (Aumann 1974). Suppose you are involved in a repeated three-player game. Then it is often reasonable to assume that the other two players correlate their choices. This might be the case, for example, if they represent two companies that form an oligopoly. In a situation like this one, you would not assume that the other players choose their strategies independently; instead, the process governing their choice of strategies will exhibit dependencies.[11]

[10] A similar kind of analogical prediction is developed in the context of exchangeable sequences by Skyrms (1993a), Festa (1997) and Romeijn (2006).

[11] In game theory, such considerations lead to the concept of a 'correlated equilibrium'.

The inductive logic of the next section is also set within de Finetti's conceptual scheme of outcomes and types. Before we dive into the details, let me note that this scheme is a departure from Carnap's framework. Since Carnap allows predicates to belong to different families, his inductive logic allows a distinction between types and outcomes. However, it operates at the level of so-called *Q*-predicates. Each *Q*-predicate is a maximally specific description of observations in terms of the underlying families of predicates (like types and outcomes). The inductive logic assigns predictive probabilities to observing a *Q*-predicate given past observations of *Q*-predicates. This is different from predictive inference based on partial exchangeability, which requires that predictive probabilities be assigned to outcomes given past observations and given their particular type; types, however, are not assigned predictive probabilities.

There certainly are ways of supplementing the probability distribution to make inferences about types possible. It would be interesting to explore how this affects analogy. This task is beyond the reach of the present paper, though.

2. A new analogical inductive system

2.1 The postulates

Let N be the total number of observations made thus far, $\boldsymbol{X}_N = (X_1, \ldots, X_N)$ the sequence of outcomes, and $\boldsymbol{Y}_N = (Y_1, \ldots, Y_N)$ the sequence of types. Furthermore, let n_{im} denote the number of times outcome i of type m is recorded in the first N trials, and let $N_m = \sum_i n_{im}$ be the number of times type m has been observed. The set of outcomes is $\{1, \ldots, s\}$, and the set of types is $\{1, \ldots, t\}$.

The following principle is a modification of Johnson's sufficientness postulate that applies to the scheme of outcomes and types:

$$(5)\quad \mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j] = f_{ij}(n_{ij}, N_j)$$

This principle says that the predictive probability of i being the next outcome, given that it is of type j, is a function of i and j as well as n_{ij} and N_j. Clearly, (5) makes analogical reasoning impossible. For example, if j are male subjects in a medical experiment and i is a particular symptom, then the number of times i has been observed in female subjects has no effect on the predictive probabilities for male subjects. In order to make room for analogy we must allow that predictive probabilities are functions of n_{ij} for *all* types j. Formally, for all $i = 1, \ldots, s$ and $j = 1, \ldots, t$,

$$(6)\quad \mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j] = f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t)$$

In the sequel we shall refer to this principle as the *modified sufficientness postulate*. Clearly, the modified sufficientness postulate should hold in all instances of analogical inference where outcomes of different types can influence each other's conditional probabilities (and no other outcomes can). This covers many of our previous examples, such as flipping two coins, observing different types of metal, observing male and female patients, or predicting the choices of two players who might correlate their strategies. Notice that our modified sufficientness postulate subsumes (5), as well as the case of exchangeability (where predictive probabilities depend on the sums $\sum_j n_{ij}$, $\sum_j N_j$).

Besides (6), it would seem reasonable to require that outcomes be partially exchangeable with respect to types. A straightforward implementation of this idea runs into problems, though. Corollary 2 below bears witness to the fact that in the presence of other plausible assumptions—including the modified sufficientness postulate (6)—partial exchangeability leaves no room for non-trivial analogical reasoning. Without going into details, I'd like to indicate the basic difficulty here in order to say how it can be overcome.

Recall that a partially exchangeable probability assignment is invariant under arbitrary permutations within a type. Thus, in the present context we say that the sequence $X_1, \ldots, X_N$ is partially exchangeable if the probability of $X_1 = x_1, \ldots, X_N = x_N, Y_1 = y_1, \ldots, Y_N = y_N$ depends only on the numbers n_{ij}. Once we fix a sequence of types, reordering outcomes within types has no effect on probabilities. This implies that

$$\begin{aligned}&\mathbb{P}[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]\\ &= \mathbb{P}[X_{N+1} = k, X_{N+2} = l, X_{N+3} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]\end{aligned}$$

for *all* types j, m, and *all* outcomes i, l, k.[12] The foregoing equation significantly constrains how type m can influence predictive

[12] This follows from the fact that the conditional probability $\mathbb{P}[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]$ is equal to

$$\frac{\mathbb{P}[\boldsymbol{X}_N, \boldsymbol{Y}_N, X_{N+1} = i, X_{N+2} = l, X_{N+3} = k, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]}{\mathbb{P}[\boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]}$$

probabilities of type j. If $l = k$, the left-hand side of the equation is a conditional probability of an event where a k-outcome of type m happens before a k-outcome of type j. The corresponding event on the right-hand side has the k-outcome of type j happen before the k-outcome of type m. For this to hold, the effect m has on j needs to exactly balance out the influence of j on m regardless of the number of outcomes and types that are involved. The statement of Corollary 2 is a precise way of saying what is intuitively plausible: that this is possible only under very special circumstances.

Let me illustrate the issue with the example of observing symptoms in female and male patients. Suppose we think that a male observation of some symptom A has a persistent influence on whether the next female patient also shows A. Then partial exchangeability would require that influence to be the same no matter when observations of A occur in the male sample. In the presence of the other axioms, the constraints this creates are too severe for the persistent analogy influence to take on a non-trivial form.

There might be more than one way to avoid this difficulty. In my view, the other axioms are reasonable, so I opt for the following modification of partial exchangeability.[13] For all types j, m, and for all outcomes i, l, k such that $i, k \neq l$,

$$(7a) \quad \begin{aligned} &\mathbb{P}[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j] \\ &= \mathbb{P}[X_{N+1} = k, X_{N+2} = l, X_{N+3} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j] \end{aligned}$$

In addition, we also postulate that

$$(7b) \quad \begin{aligned} &\mathbb{P}[X_{N+1} = i, X_{N+2} = k | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = j] \\ &= \mathbb{P}[X_{N+1} = k, X_{N+2} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = j] \end{aligned}$$

for all types j and all outcomes i, k.[14] Taken together, I refer to (7a) and (7b) as *weak partial exchangeability*. The first part says that the conditional probabilities in (7a) may differ if $k, i = l$, which is more

By partial exchangeability, the numerator is equal to $\mathbb{P}[\boldsymbol{X}_N, \boldsymbol{Y}_N, X_{N+1} = k, X_{N+2} = l, X_{N+3} = i, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]$, which yields the right-hand side of the equation.

[13] An alternative would be to keep partial exchangeability but modify other postulates in order to build appropriate dependencies between types into the chance prior of the de Finetti representation, along the lines of Skyrms (1993a), Festa (1997) or Romeijn (2006).

[14] Conditions (7a) and (7b) could be defined equivalently in terms of unconditional probabilities which exhibit order invariance for all outcomes i, j within a given type unless there is an intermediate outcome i or j of another type.

liberal than partial exchangeability requires, and avoids the difficulty explained above. At the same time, (7a) and (7b) preserve much of partial exchangeability by insisting that reordering outcomes never affects conditional probabilities whenever there are no, or at least no relevant, intermediate outcomes.

We will need two axioms besides the modified sufficientness postulate (6) and weak partial exchangebility (7). The following regularity condition guarantees that all conditional probabilities are well-defined:

$$(8)\quad \mathbb{P}[X_1 = x_1, \ldots, X_{N+1} = x_{N+1}, Y_1 = y_1, \ldots, Y_{N+1} = y_{N+1}] > 0$$

for all combinations of outcomes $x_1, \ldots, x_{N+1}$ and for all combinations of types $y_1, \ldots, y_{N+1}$. Our final assumption requires that outcomes and types are to some extent independent. For all outcomes i and all types j, k, l,

$$(9)\quad \begin{aligned} &\mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j] \\ &= \mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = k] \\ &= \mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = k, Y_{N+3} = l] \end{aligned}$$

This says that X_{N+1} and Y_{N+2} are conditionally independent given $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_{N+1}$, and that X_{N+1} and Y_{N+3} are conditionally independent given $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_{N+2}$.[15] If we think of types as different kinds of experiment, then the experiments planned for trials $N + 2$ and $N + 3$ are independent of the outcome of trial $N + 1$. Condition (9) is thus a substantive assumption. To see how it might fail, suppose we can choose between different medical treatments; if one of them leads to a success tomorrow, we might use it with higher probability on the next occasion. In this case, the choice of future treatments depends on tomorrow's outcome. On the other hand, postulate (9) is adequate whenever types of experiments are chosen at random (independently of outcomes).

2.2 *The main result*

Our main theorem says that the four axioms of the previous section give rise to a parametrized family of analogical predictive probabilities.

[15] In the formal development, the postulate allows us to apply the sufficientness postulate to predictive probabilities that condition on future types, such as those in (9). See the appendix.

The parameters fall into two groups: the first group expresses initial beliefs (as the alpha parameters in the Johnson-Carnap continuum of inductive methods); the second group consists of analogy parameters.

Theorem 1. Let $X_1, \ldots, X_{N^*}$ $(N^* \geq 4)$ be a sequence of random variables that take on values in $\{1, \ldots, s\}$, $3 \leq s \leq \infty$. Let $Y_1, \ldots, Y_{N^*}$ be a sequence of types taking values in $\{1, \ldots, t\}$, $1 \leq t < \infty$. Suppose that for every $N < N^*$, the four postulates (6), (7), (8) and (9) hold. Let $\sum_i n_{ij} = N_j$ and $\sum_j N_j = N$. Suppose, moreover, that for each type j, outcomes of that type are not independent of each other. Then there exist non-zero constants κ_{ij} (for each j, either all κ_{ij} are positive or all are negative) and constants β_{jm}, $1 \leq i \leq s$, $1 \leq j, m \leq t$, $m \neq j$, such that $N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j \neq 0$ (where $K_j = \sum_i \kappa_{ij}$) and

$$(10)\quad \mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j] = \frac{n_{ij} + \sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}$$

for all $N < N^*$ and all $0 \leq n_{ij} \leq N_j$, $1 \leq i \leq s$, $1 \leq j \leq t$.

Remark 1. It is well known that Johnson's sufficientness postulate is vacuous in Carnapian inductive logic if the number of outcomes $s = 2$. This is also true of the modified sufficientness postulate (6). In addition, the application of (7a) requires there to be at least three distinct outcomes (see Lemma 2 and Lemma 3 in the appendix); this is because (7a) also is vacuous if there are only two types. A part of this difficulty could be overcome by assuming that predictive probabilities are linear in the numbers n_{ij} (see Lemma 1). Overall, an alternative approach could probably be developed with the help of relevance quotients, as suggested by Costantini (1979), but at the moment I don't see how this would play out in detail.

Remark 2. Theorem 1 provides predictive probabilities only for outcomes, and not for types. Hence, as in the case of partially exchangeable probability distributions, those predictive probabilities do not generate a probability distribution over the full process of outcomes *and* types. This makes it difficult to say for which values of the parameters in (10) the postulates (6), (7), (8) and (9) hold. What we can say is that, in the presence of (8) and (9), the first two postulates (6) and (7) hold for all parameters κ_{ij}, β_{jm} as specified in the theorem. We cannot say, though, whether (7) is necessary for the representation (10). Theorem 1 only provides sufficient conditions for the inductive method (10).

Before saying more about Theorem 1, I would like to note two immediate consequences. First, the parameters of the inductive method (10) are more constrained if $X_1, X_2, \ldots$ is an infinite sequence.

Corollary 1. Let $X_1, X_2, \ldots$ be an infinite sequence of outcomes such that the assumptions of Theorem (1) hold. Then $\kappa_{ij} > 0$ and $\beta_{jm} \geq 0$, $1 \leq i \leq s$, $1 \leq j, m \leq t, m \neq j$.

The second corollary shows the consequences of assuming partial exchangeability instead of weak partial exchangeability. If (7a) holds without restrictions, then either types must be independent or outcomes must be exchangeable across types.

Corollary 2. Let $X_1, X_2, \ldots X_{N^*}$ ($N^* \geq 4$) be a sequence of outcomes such that the assumptions of Theorem 1 hold. Then (7a) holds for l = k if and only if either $\beta_{jm} = \beta_{mj} = 0$ or $\beta_{jm} = \beta_{mj} = 1$.

Hence, if (7a) holds for all outcomes, there are no analogy influences short of full analogy; the inductive system basically reduces to the Johnson-Carnap continuum of inductive methods.

The proofs of Theorem 1 and its corollaries can be found in the appendix. The proof of Theorem 1 proceeds in two steps. The first step is to show that predictive probabilities of $X_{N+1} = i$ given $Y_{N+1} = j$ are linear functions of $n_{i1}, \ldots, n_{it}$:

$$a_{ij} + b_{j1} n_{i1} + \ldots + b_{jt} n_{it}$$

The parameters $a_{ij}, b_{j1}, \ldots, b_{jt}$ depend on $N_1, \ldots, N_t$, but not on the specific numbers of outcomes. This step is based on the modified sufficientness postulate and the large system of linear equations it gives rise to.

By normalizing the parameters $a_{ij}, b_{j1}, \ldots, b_{jt}$, predictive probabilities can be brought into the form required by (10). The second step of the proof shows that the normalized parameters, $\kappa_{ij}, \beta_{j1}, \ldots, \beta_{jt}$, do not depend on $N_1, \ldots, N_t$. This step relies on weak partial exchangeability (7) and the fact that the parameters are the same for all possible patterns of outcomes that add up to $N_1, \ldots, N_t$.

The predictive probabilities of an inductive system can often be generated by an urn model. Even if it is not known whether a set of axioms is sufficient for an inductive system, the existence of such an urn model proves the joint consistency of the axioms. For instance, the Johnson-Carnap continuum of inductive methods is mathematically equivalent to *Polya urn processes*. A Polya urn process starts with an urn containing a finite number of balls of different colours. After choosing a ball at random, it is returned to the urn with another ball of the same colour before the process is repeated.[16]

Analogical predictive probabilities can also be generated by urn processes. Let's start with the case $N^* = \infty$. First, choose an infinite sequence of types $(y_1, y_2, \ldots)$ according to some probability distribution that assigns positive probability to every finite initial sequence of types. For each type j, there is an urn containing κ_{ij} balls of colour i (one colour for each outcome). The process starts with choosing a ball at random from urn y_1, observing the colour of the ball, and replacing it with two balls of the same colour. In addition, $\beta_{jm} \geq 0$ balls of that colour are put in every urn m. This procedure is repeated for urns $y_2, y_3, \ldots$ The same urn model can be used if $N^* < \infty$, as long as all parameters are non-negative.

Consider now the case where all κ_{ij} are negative and all β_{jm} are such that $|\beta_{jm}| \leq 1$. Suppose that $N^* < min\{-\kappa_{ij}\}$. Negative κ_{ij} correspond to choosing a ball from an urn without replacement, since there is an upper bound to the number of observations n_{ij} we can make before the predictive probabilities (10) become negative. We suppose that for each type j there is an urn containing $-\kappa_{ij} > 0$ balls of colour i. We start by choosing a finite sequence of types of length N^* from a distribution that assigns every such sequence a positive probability. Tracking the sequence of types, balls are then chosen without replacement from urns. If a ball of colour i is chosen from an urn of type j, then it is not replaced and β_{jm} balls of colour i are taken out of urn m (or added, in case β_{jm} is negative).

2.3 Some consequences

A closer look at the beta parameters reveals the sense in which they represent analogy effects. Each β_{jk} depends both on the prior κ_{ij} and on the conditional probability $p_{ij,\,ik} = \mathbb{P}[X_2 = i | X_1 = i, Y_1 = k, Y_2 = j]$. More precisely, since

$$p_{ij,\,ik} = \frac{\beta_{jk} + \kappa_{ij}}{\beta_{jk} + K_j}$$

[16] Sampling balls without replacement also corresponds to certain Carnapian systems. For more information see Kuipers (1978).

β_{jk} can be written as

$$(11) \quad \beta_{jk} = \frac{p_{ij,ik}K_j - \kappa_{ij}}{1 - p_{ij,ik}}$$

By assumption, $p_{ij,ik} < 1$; moreover, $\mathbb{P}[X_1 = i | Y_1 = j] = \kappa_{ij}/K_j$. This implies that β_{jk} is positive, negative or 0, according to whether

$$p_{ij,ik} > \mathbb{P}[X_1 = i | Y_1 = j], \quad p_{ij,ik} < \mathbb{P}[X_1 = i | Y_1 = j], \quad p_{ij,ik} = \mathbb{P}[X_1 = i | Y_1 = j]$$

respectively. Thus the sign of β_{jk} depends on whether observing an outcome–type pair (i, k) increases the probability of observing a pair (i, j) beyond its prior probability. The absolute value of β_{jk} increases as the probability $p_{ij,ik}$ gets closer to 1; the larger the effect of an (i, k) observation on the conditional probability, the higher the value of β_{jk}. So β_{jk} is a probabilistic measure of the influence of observations of type k on observations of type j. According to the inductive method (10), this influence is invariant with regard to the observational context: no combination of additional observations changes the way k types affect the conditional probabilities of j types. Thus β_{jk} can be thought of as the fixed analogy influence type k exerts on type j.

The quantities β_{jk} meet a natural analogy requirement: if $\beta_{jk} \geq \beta_{jl}$, then the influence of k on j is larger than the influence of l on j. This is made precise in the following proposition.[17]

Proposition 1. Suppose that all assumptions of Theorem 1 hold, and that for type j, all κ_{ij} are positive. If $\beta_{jk} \geq \beta_{jl}$, then for all $N < N^*$,

$$(12) \quad \begin{aligned} &\mathbb{P}[X_{N+1} = i | \boldsymbol{X}_{N-1}, \boldsymbol{Y}_{N-1}, X_N = i, Y_N = k, Y_{N+1} = j] \\ &\geq \mathbb{P}[X_{N+1} = i | \boldsymbol{X}_{N-1}, \boldsymbol{Y}_{N-1}, X_N = i, Y_N = l, Y_{N+1} = j] \end{aligned}$$

If $\beta_{jk} > \beta_{jl}$, then the inequality is strict.

Proof. By Theorem 1, (12) is equivalent to

$$\frac{n_{ij} + \beta_{jk}(n_{ik} + 1) + \sum_{m \neq j,k} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \beta_{jk}(N_k + 1) + \sum_{m \neq j,k} \beta_{jm} N_m + K_j} \geq \frac{n_{ij} + \beta_{jl}(n_{il} + 1) + \sum_{m \neq j,l} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \beta_{jl}(N_l + 1) + \sum_{m \neq j,l} \beta_{jm} N_m + K_j}$$

[17] The proposition does not consider the case corresponding to sampling from urns without replacement, which is somewhat less intuitive, but can be treated in a similar way.

where the n_{ij}, N_j, etc. denote counts up to period $N-1$. Rearranging and simplifying shows that this inequality is equivalent to

$$(13)\quad (\beta_{jk}-\beta_{jl})\Big(K_j-\kappa_{ij}+N_j+\sum\nolimits_{m\neq j}\beta_{jm}N_m-\Big(n_{ij}+\sum\nolimits_{m\neq j}\beta_{jm}n_{im}\Big)\Big)\geq 0$$

Since

$$N_j+\sum\nolimits_{m\neq j}\beta_{jm}N_m\geq n_{ij}+\sum\nolimits_{m\neq j}\beta_{jm}n_{im}$$

and because $\beta_{jk}\geq\beta_{jl}$, the inequality (13) holds if

$$K_j-\kappa_{ij}\geq 0$$

Since the parameters κ_{ij} are positive, the assertion of the proposition follows. ■

Proposition 1 says that the larger β_{jk}, the more influence observations of type k have on the predictive probabilities for type j.

Theorem 1 allows $\beta_{jk}>1$; in this case, observations of type k have a larger influence on the predictive probabilities for type j than observations of type j itself. This might be reasonable in situations where observations of type k are considered to be more reliable. However, analogy is often thought of in terms of similarity. Since each type is most similar to itself, other types should not exhibit a larger analogy effect on that type than it has on itself. This idea can easily be incorporated into the present framework by requiring that, in addition to the axioms assumed by Theorem 1, the following assumption holds: for all $k\neq j$,

$$(14)\quad \mathbb{P}[X_2=i\,|\,X_1=i,Y_1=j,Y_2=j]\geq\mathbb{P}[X_2=i\,|\,X_1=i,Y_1=k,Y_2=j]$$

This says that an (i, j) observation's influence on an (i, j) prediction cannot be less than the influence an (i, k) observation has on an (i, j) prediction. It follows that every type has maximal predictive influence on itself.

Proposition 2. Suppose that all assumptions of Theorem 1 hold, and that $\kappa_{ij}>0$ for all outcomes i of type j. If (14) is true for all $k\neq j$, then $\beta_{jk}\leq 1, k\neq j$.

Proof. Suppose that (14) holds. It follows from Theorem 1 that

$$\frac{1+\kappa_{ij}}{1+K_j}\geq\frac{\beta_{jk}+\kappa_{ij}}{\beta_{jk}+K_j}$$

This implies

$$(1-\beta_{jk})(K_j-\kappa_{ij})\geq 0$$

Since all κ_{ij} are positive, it follows that $\beta_{jk}\leq 1$. ■

In conjunction with Proposition 1, this implies

$$\mathbb{P}[X_{N+1}=i|\boldsymbol{X}_{N-1},\boldsymbol{Y}_{N-1},X_N=i,Y_N=j,Y_{N+1}=j]$$
$$\geq \mathbb{P}[X_{N+1}=i|\boldsymbol{X}_{N-1},\boldsymbol{Y}_{N-1},X_N=i,Y_N=k,Y_{N+1}=j]$$

for all $k\neq j$.

Considering analogy in terms of similarity leads to a number of important special cases. If the number of types is equal to 1, the theorem reduces to the Johnson-Carnap continuum of inductive methods. But the same happens if all $\beta_{jm}=1$; in this case types are indistinguishable. If each type is maximally similar to itself, but all other types are equally similar to it, we have $\beta_{jm}=\beta\neq 1$ for all j, m. This leads to a simplification of the inductive rule (10):

$$\mathbb{P}[X_{N+1}=i|\boldsymbol{X}_N,\boldsymbol{Y}_N,Y_{N+1}=j]=\frac{n_{ij}+\beta(N^i-n_{ij})+\kappa_{ij}}{N_j+\beta(N-N_j)+K_j}$$

where $N^i=\sum_m n_{im}$ is the number of outcomes i regardless of type.

Similarity is usually thought to be a symmetric relationship. In our framework this can be expressed by requiring that $\beta_{jm}=\beta_{mj}$. No similarity means that $\beta_{jm}=0$. If this holds for all types $m\neq j$, the new inductive rule again reduces to the Johnson-Carnap continuum, since inductive probabilities can then be analysed irrespective of other types. By Corollary 2, this case corresponds to assuming partial exchangeability within our framework.

3. Reichenbach's axiom

3.1 Asymptotic properties

Perhaps the most controversial aspect of our new inductive logic is an apparent violation of *Reichenbach's axiom* (also known as the *axiom of convergence*). Reichenbach's axiom is a well-known principle in inductive logic. It requires that predictive probabilities converge to sample frequencies in the limit, provided that they exist. The main motivation for Reichenbach's axiom is the very reasonable idea that information about the sample should outweigh our initial opinions as we get more information.

The Johnson-Carnap continuum of inductive methods satisfies Reichenbach's axiom. Suppose the limiting relative frequency of outcome i, p_i, exists with probability one (as is the case for exchangeable probability distributions). Then the predictive probabilities

$$\frac{n_i + \alpha_i}{n + \sum_j \alpha_j}$$

converge to p_i with probability 1; in particular, the alpha parameters are outweighed in the long run by information from the sample.

The inductive system of this paper leads to a more complex situation. Let's consider the limit of analogical predictive probabilities (10) for infinite sequences $X_1, X_2, \ldots$ Suppose the limit of $\rho_{jk} = N_j/N_k$ exists for all pairs of types j, k. Let A_{jk} be given by

$$A_{jk} = \frac{\beta_{jk} N_k}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}$$

and let B_j be given by

$$B_j = \frac{K_j}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}$$

With this notation, predictive probabilities can be expressed as follows:

$$\begin{aligned} &\mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j] \\ &= \left(1 - \sum_{k \neq j} A_{jk} - B_j\right) \frac{n_{ij}}{N_j} + \sum_{k \neq j} A_{jk} \frac{n_{ik}}{N_k} + B_j \frac{\kappa_{ij}}{K_j} \end{aligned}$$

Suppose now that each type occurs infinitely often, and that $lim\, n_{ij}/N_j = \eta_{ij}$ exists for all outcomes i and all types j. Then predictive probabilities converge to

$$\left(1 - \sum_{k \neq j} \frac{\beta_{jk}}{\rho_{jk} + \sum_{m \neq j} \beta_{jm} \rho_{mk}}\right) \eta_{ij} + \sum_{k \neq j} \frac{\beta_{jk}}{\rho_{jk} + \sum_{m \neq j} \beta_{jm} \rho_{mk}} \eta_{ik}$$

As a consequence, the predictive probability $\mathbb{P}[X_{N+1} = i | X_N, \boldsymbol{Y}_N, \boldsymbol{Y}_{N+1} = j]$ generally does not converge to the limiting relative frequency η_{ij}; instead, its limit is a convex combination of relative frequencies η_{ik}, $1 \leq k \leq t$. Let's suppose, for example, that all types are observed approximately equally often: $\rho_{jk} = 1$, $1 \leq j, k \leq t$, $k \neq j$. Then $\mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j]$ is given by

$$(15) \quad \frac{1}{1 + \sum_{k \neq j} \beta_{jk}} \left(\eta_{ij} + \sum_{k \neq j} \beta_{jk} \eta_{ik} \right)$$

Unless $\beta_{jk} = 0$ for all $k \neq j$ or $\eta_{ik} = \eta_{ij}$ for all $k \neq j$, this expression is not equal to η_{ij}. Now, one reading of Reichenbach's axiom is that $\mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j]$ needs to converge to η_{ij}. Our inductive system clearly violates this version of Reichenbach's axiom.

In my view, this conclusion fails to appreciate an important point about Reichenbach's axiom. If Reichenbach's axiom really requires that $\mathbb{P}[X_{N+1} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j]$ converges to η_{ij}, then analogy can *only* be part of one's prior opinion, which fades as one's information of the sample grows without bound. This certainly is an important sense in which analogy plays a role in inductive inferences. It underlies de Finetti's model of partial exchangeability, and it is the type of analogy captured, in different ways, by the inductive systems of Spohn (1981), Kuipers (1984), Skyrms (1993a), Festa (1997), Maher (2000, 2001), Romeijn (2006), or Hill and Paris (2013). The basic idea can be understood most easily for systems that have a de Finetti representation, such as de Finetti's own model of partial exchangeability or models that preserve exchangeability. In these models, analogy is expressed through the *chance prior*, and not through the *chance set-up*. The chance prior allows correlations between outcomes or types, but the chance set-up consists of independent trials that obliterate all correlations in the long run.[18] The reason is that sequences of outcomes are assumed to be *conditionally independent* on the chances of outcomes; once chances are fixed, the process unfolds without any dependencies among the outcomes. Since the process is fully captured by chances of outcomes, predictive probabilities ought to discover chances in the long run: the conditional probability of outcome *i* given past observations should get closer to the chance of *i*. Chances almost surely coincide with limiting relative frequencies. This is why Reichenbach's axiom, in the above version, is very plausible for processes that are conditionally independent given chances of outcomes.

But chance processes need not be conditionally independent given chances of outcomes. They can exhibit a more complex structure of

[18] I am indebted to Jan-Willem Romeijn for suggesting to think about these issues in terms of chance priors and chance set-ups.

probabilistic dependencies. This is arguably the case for the chance processes associated with our inductive system. Unlike partial exchangeability, weak partial exchangeability (7) explicitly allows probabilities to be sensitive to some ways of reordering outcomes within a type. This suggests that the underlying chance set-up must allow for permanent correlations between types, which can be illustrated by the example of clinical trials with female and male patients, or with the possibility of correlated strategy choice in games.

At the moment I do not know more about the structure of the relevant chance set-up. A fully satisfying answer would involve proving a de Finetti representation theorem for weakly partially exchangeable probability assignments. This would provide us with a new perspective on understanding the inductive system (10) besides the axiomatic approach of this paper.[19] Exploring this issue is a topic for future research. What I hope is that my qualitative remarks are sufficient to reject the above version of Reichenbach's axiom as the uniquely plausible one. If outcomes of distinct types are dependent according to the chance set-up associated with our inductive system, then the predictive probabilities of an outcome should not converge to the limiting relative frequency of that outcome within its type; for this would imply that predictive probabilities ignore crucial information about the process.

Similar remarks apply to the inductive systems of Niiniluoto (1981) and Costantini (1983), which also appear to violate Reichenbach's axiom.[20] Niiniluoto's and Costantini's systems operate at the level of Q-predicates. Neither of the systems gives rise to exchangeable probability assignments. This indicates that their underlying chance processes are not independent trials. But if there are persistent dependencies between outcomes, they should be reflected in the limiting behaviour of predictive probabilities, which would explain why they violate Reichenbach's axiom.

Summing up, violations of Reichenbach's axiom should not automatically be regarded as good arguments against inductive systems. Reichenbach's axiom implicitly assumes that only information about sample frequencies should matter in the long run. This is appropriate for settings where observations of outcomes are essentially independent. But, at least in its most basic form, it fails to be an adequate

[19] See Romeijn (2006) for a discussion of chance models and their virtues in inductive logic.

[20] Spohn (1981) discusses Reichenbach's axiom in the context of Niiniluoto's system.

requirement in situations where types depend on each other in more complex ways.

3.2 Direct and indirect analogy

The preceding discussion suggests distinguishing analogical inductive inference according to whether analogy effects are *direct* or *indirect*. An inductive system exhibits indirect analogy effects if observations of an outcome have an impact on the predictive probabilities of another outcome only through prior opinions, as in the models of Skyrms (1993a), Festa (1997), Romeijn (2006), and Hill and Paris (2013). In our model, as in the models by Niiniluoto (1981) and Costantini (1983), analogy judgements have a direct effect on predictive probabilities via how conditional probabilities are updated. Which type of analogy is appropriate depends on the details of the epistemic situation. If covariates between types are thought to be persistent, a model of direct analogy is called for; otherwise, a model of indirect analogy should be chosen. Mixtures of the two types are also conceivable.

It would be desirable to extend our inductive system so as to encompass both direct and indirect analogy. The most straightforward way to do this would be to allow the beta parameters to vary, but this approach faces severe difficulties. The modified sufficientness postulate (6), together with regularity (8), implies the rather strong conclusion that the analogy parameters b_{jm} in (16) only depend on $N_1, \ldots, N_t$ (see Lemma 1 below). This is a significant constraint on predictive probabilities. For instance, if analogy influences ought to decrease in $N_1, \ldots, N_t$, the symmetries of the inductive system are bound to get considerably more complex than weak partial exchangeability. Suppose, for instance, that analogy parameters are given by $1/N$. Then the conditional probability of observing outcomes i and k (both of type j) at trials $N+1, N+2$ is

$$\frac{n_{ij} + \frac{1}{N}\sum_{m\neq j} n_{im} + \kappa_{ij}}{N_j + \frac{N-N_j}{N} + K_j} \frac{n_{kj} + \frac{1}{N+1}\sum_{m\neq j} n_{km} + \kappa_{kj}}{N_j + 1 + \frac{N-N_j}{N+1} + K_j}$$

It is not difficult to see that j and k cannot in general be switched without affecting the conditional probability; but this violates (7b).

A more promising alternative is to embed our inductive logic in a hierarchical model with a prior probability measure over the beta parameters. This means that analogy effects are modelled as unknown quantities that are subject to uncertainty. In such a higher-order model, the value of beta parameters can also change in the light of evidence, which is attractive because it allows learning from experience about analogies. Furthermore, certain priors might lead to beta parameters that vanish in the limit. As a result, hierarchical models may successfully unify direct and indirect analogy effects.

3.3 Finite sequences

Let me end this section by drawing attention to a domain of inductive reasoning that falls outside the purview of Reichenbach's axiom. Often, a finite model might be more appropriate than an infinite one, for example, if we don't plan to make a large, unbounded number of observations. For a finite horizon, the fact that predictive probabilities don't always converge to limiting relative frequencies appears to be of little consequence. What matters in the finite setting is whether predictive probabilities reflect all the relevant information, including analogical relationships. As long as we think that analogy effects are approximately constant during a finite experiment, even though they may vanish in the limit, our inductive system is applicable.

As an illustration consider again Achinstein's example of investigating whether different types of metal conduct electricity. In this case it is natural to think that predictive probabilities should converge to limiting relative frequencies. But if the experiment only consists of a finite number of trials, our inductive system might serve as a useful model, either because constant analogy effects are compatible with our prior knowledge over a finite period of time or because analogy effects can be captured approximately by constant parameters.

4. Concluding remarks

As we observed in §1, Carnapian inductive logic fails to capture important types of analogical inference for two reasons. One is that exchangeability assumes 'too much' analogy, since trials are assumed to be perfectly similar. The other reason is that Johnson's sufficientness postulate amounts to assuming 'too little' analogy by making it impossible that observations of an outcome exert some influence on the predictive probabilities of similar outcomes. This provides us with two levers for altering analogy judgements, both of which have been pulled in the inductive logic literature. Some inductive systems abandon

Johnson's sufficientness postulate but preserve exchangeability (Skyrms 1993a; Festa 1997; Romeijn 2006; Hill and Paris 2013); others abandon exchangeability (de Finetti 1938); while systems like ours give up on both exchangeability and Johnson's sufficientness postulate.

I do not wish to suggest that my model is the only one that adequately captures analogical inductive reasoning. Such a claim would be foolish in the face of the many different forms of analogical reasoning we use. In my view, many of the inductive systems I have mentioned above have a domain of applicability. Instead of searching for the uniquely rational way to reason analogically, I think it is much more valuable to heed W. E. Johnson's advice:

> The postulate adopted in a controversial kind of theorem cannot be generalised to cover all sorts of working problems; so that it is the logician's business, having once formulated a specific postulate, to indicate very carefully the factual or epistemic conditions under which it has practical value. (Johnson 1932, pp. 418–19)

When trying to gauge the adequacy of an inductive system, we should determine the principles and assumptions from which it can be derived. The axiomatic approach of Johnson and Carnap achieves this goal by formulating axioms about the predictively relevant aspects of the observational process. De Finetti's approach identifies the chance set-up associated with a prior distribution. The two approaches are complementary ways of conveying the epistemic conditions under which an inductive system should be adopted. In this paper I have followed Johnson's and Carnap's methodology by identifying the local symmetries associated with an important class of analogical inferences. De Finetti's approach, which is left as an open problem, would provide us with a more complete understanding of the new inductive system.

There are other open problems. Let me mention one that is particularly important for analogical reasoning. It would be desirable to extend analogical inductive systems to real-valued random variables. Because the domain of these random variables has a metric, there often is a natural sense of analogy or similarity. Zabell (2011) discusses analogical inference in this setting based on a number of well-known symmetry assumptions. The inductive logic introduced in Skyrms (1993b), which is based on the Blackwell-MacQueen urn process (Blackwell and MacQueen 1973), could be used as a starting point for the analysis of different types of analogical reasoning in continuum probability spaces.[21]

Appendix: Proof of Theorem 1 and its corollaries

The proof of Theorem 1 is an adaptation of Johnson's argument (Johnson 1932; Zabell 1982) to the present setting. In fact, if the number of types $t = 1$, the argument reduces to this proof. We thus assume throughout that the number of types t is at least 2.

We start by showing that our analogue of the sufficientness postulate implies that predictive probabilities are linear in $n_{i1}, \ldots, n_{it}$.

Lemma 1 . If (6) and (8) hold, then for every outcome i and every type j there exist constants $a_{ij} > 0$ and $b_{j1}, \ldots, b_{jt}$ such that

$$(16) \quad f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = a_{ij} + b_{j1}n_{i1} + \ldots + b_{jt}n_{it}$$

where the constants depend only on $N_1, \ldots, N_t$. Furthermore, the constants $b_{j1}, \ldots, b_{jt}$ do not depend on the outcome i.

Proof. Fix an outcome i and a type j. We start with considering the influence of type 1 outcomes. Suppose that $N_1 \geq 2$, and that $0 < n_{i1}, n_{k1}$ and $n_{i1}, n_{l1} < N_1$, where i, k, l are distinct outcomes. Let

$$\begin{aligned}
\boldsymbol{N}_1 &= (n_{11}, \ldots, n_{i1}, \ldots, n_{k1}, \ldots, n_{l1}, \ldots, n_{s1}) \\
\boldsymbol{N}_2 &= (n_{11}, \ldots, n_{i1} + 1, \ldots, n_{k1} - 1, \ldots, n_{l1}, \ldots, n_{s1}) \\
\boldsymbol{N}_3 &= (n_{11}, \ldots, n_{i1}, \ldots, n_{k1} - 1, \ldots, n_{l1} + 1, \ldots, n_{s1}) \\
\boldsymbol{N}_4 &= (n_{11}, \ldots, n_{i1} - 1, \ldots, n_{k1}, \ldots, n_{l1} + 1, \ldots, n_{s1})
\end{aligned}$$

Notice that these vectors have the same number N_1 of outcomes of type 1. Using the values given in $\boldsymbol{N}_1, \boldsymbol{N}_2, \boldsymbol{N}_3, \boldsymbol{N}_4$, the equalities

$$\sum_r f_{rj}(n_{r1}, \ldots, n_{rt}) = 1$$

hold for any $n_{rm}, r = 1, \ldots, s,\ m = 2, \ldots, t$ such that $\sum_r n_{rm} = N_m$. We suppress here the counts $N_1, \ldots, N_t$, which are held fixed, in

[21] I would like to thank Hannes Leitgeb, Jeff Paris, Jan-Willem Romeijn, Brian Skyrms, Marta Sznaijder, Theo Kuipers, Sandy Zabell, and the referees and editors at *Mind* for helpful comments, as well as the audiences at the MCMP Munich, the workshop on Topics in Inductive Logic at UC Irvine, an the TARK 2015 conference at Carnegie Mellon University.

order to simplify notation. This results in four equations, for example the following two, which correspond to $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$:

$$f_{ij}(n_{i1}, \ldots, n_{it}) + f_{kj}(n_{k1}, \ldots, n_{kt}) + \sum\nolimits_{r \neq i, k} f_{rj}(n_{r1}, \ldots, n_{rt}) = 1$$

$$f_{ij}(n_{i1} + 1, \ldots, n_{it}) + f_{kj}(n_{k1} - 1, \ldots, n_{kt}) + \sum\nolimits_{r \neq i, k} f_{rj}(n_{r1}, \ldots, n_{rt}) = 1$$

As in the first part of the proof of Lemma 2.1 in Zabell (1982), the four equations together imply

$$\begin{aligned} & f_{ij}(n_{i1} + 1, \ldots, n_{it}) - f_{ij}(n_{i1}, \ldots, n_{it}) \\ & \quad = f_{kj}(n_{k1}, \ldots, n_{kt}) - f_{kj}(n_{k1} - 1, \ldots, n_{kt}) \\ & \quad = f_{lj}(n_{l1} + 1, \ldots, n_{lt}) - f_{lj}(n_{l1}, \ldots, n_{lt}) \\ & \quad = f_{ij}(n_{i1}, \ldots, n_{it}) - f_{ij}(n_{i1} - 1, \ldots, n_{it}) \end{aligned} \tag{17}$$

(The first equality follows from the foregoing two equations.) Hence,

$$f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = f_{ij}(0, n_{i2}, \ldots, n_{it}, N_1, \ldots, N_t) + b_{j1} n_{i1}$$

where

$$b_{j1} = f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) - f_{ij}(n_{i1} - 1, \ldots, n_{it}, N_1, \ldots, N_t)$$

The second part of Zabell's proof of Lemma 2.1 shows that this also holds if $N_1 = 1$. Furthermore, b_{j1} does not depend on the choice of n_{im}, $m = 2, \ldots, t$. To see this, note that the equalities in (17) hold for all choices of n_{rm}, $r = 1, \ldots, s$, $m = 2, \ldots, t$; in particular, they hold when $n_{lm} = 0$, $m = 2, \ldots, t$ and $n_{km} = N_m - n_{im}$, $m = 2, \ldots, t$. It follows that

$$\begin{aligned} & f_{ij}(n_{i1} + 1, n_{i2}, \ldots, n_{it}) - f_{ij}(n_{i1}, n_{i2}, \ldots, n_{it}) \\ & \quad = f_{lj}(n_{l1} + 1, 0, \ldots, 0) - f_{lj}(n_{l1}, 0, \ldots, 0) \end{aligned}$$

for all n_{im}, $m = 2, \ldots, t$, and so b_{j1} does not depend on these values. Because of (17), b_{j1} also is independent of i. Therefore, b_{j1} only depends on the counts $N_1, \ldots, N_t$.

The same arguments apply to n_{i2}. The result is that

$$f_{ij}(0, n_{i2}, \ldots, n_{it}, N_1, \ldots, N_t) = f_{ij}(0, 0, n_{i3}, \ldots, n_{it}, N_1, \ldots, N_t) + b_{j2} n_{i2}$$

for some constant b_{i2} which only depends on $N_1, \ldots, N_t$. Since $t < \infty$, repeated applications of these arguments lead to

$$f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = f_{ij}(0, \ldots, 0, N_1, \ldots, N_t) + b_{j1}n_{i1} + \ldots + b_{jt}n_{it}$$

By setting

$$a_{ij} = f_{ij}(0, \ldots, 0, N_1, \ldots, N_t)$$

(16) is established. It follows from (8) that $a_{ij} > 0$. ■

Lemma 1 establishes the basic representation of our predictive probabilities. In order to bring them into the form given in (10), let $A_j = \sum_r a_{rj}$. Then, by the law of total probability,

$$\begin{aligned} A_j + b_{j1}N_1 + \ldots, + b_{jt}N_t &= \textstyle\sum_r a_{rj} + b_{j1}n_{r1} + \ldots, + b_{jt}n_{rt} \\ &= \textstyle\sum_r f_{rj}(n_{r1}, \ldots, n_{rt}) = 1 \end{aligned}$$

Suppose that $b_{jj} \neq 0$ and let $\kappa_{ij} = a_{ij}/b_{jj}$, $K_j = \sum_r \kappa_{rj} = A_j/b_{jj}$, $\beta_{jm} = b_{jm}/b_{jj}, k \neq j$. Then

$$\frac{1}{b_{jj}} = N_j + \sum\nolimits_{m \neq j} \beta_{jm} N_m + K_j$$

It follows from Lemma 1 that

$$(18) \quad f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = \frac{n_{ij} + \sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}$$

Note that a_{ij} and b_{jm} depend in general on $N_1, \ldots, N_t$. Thus it remains to show that κ_{ij}, β_{jm} do not depend on $N_1, \ldots, N_t$. This is achieved in Lemma 3 for the case where observations within a type are not independent, which is the more typical case of inductive learning. The case of independence is considered in the next lemma.

Lemma 2. Let $X_1, \ldots, X_{N+1}, X_{N+2}, X_{N+3}$, $N \geq 1$, be a sequence of outcomes and $Y_1, \ldots, Y_{N+1}, Y_{N+2}, Y_{N+3}$, $N \geq 1$, a sequence of types such that (6), (7), (8) and (9) hold. Then for all types j and m, if

$$b_{jj}(N_1, \ldots, N_m, \ldots, N_t) \cdot b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_t) = 0$$

then

$$b_{jj}(N_1, \ldots, N_m, \ldots, N_t) = b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_t) = 0$$

provided that $N_j > 0$.

Proof . Let $i \neq k$. Consider first the case $j = m$. Postulate (7b) says that

$$\begin{aligned}(19)\quad &\mathbb{P}[X_{N+1} = i, X_{N+2} = k | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = j] \\ &= \mathbb{P}[X_{N+1} = k, X_{N+2} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = j]\end{aligned}$$

Let $a_{ij} = a_{ij}(N_1, \ldots, N_j, \ldots, N_t)$, $a'_{ij} = a_{ij}(N_1, \ldots, N_j + 1, \ldots, N_t)$, $b_{jj} = b_{jj}(N_1, \ldots, N_j, \ldots, N_t)$, $b'_{jj} = b'_{jj}(N_1, \ldots, N_j + 1, \ldots, N_t)$, etc. Let $\sum_r = \sum_{m \neq j} b_{jm} n_{rm}$ and $\sum'_r = \sum_{m \neq j} b'_{jm} n_{rm}$. Then, by using (9) and (16), (19) implies

$$\begin{aligned}(20)\quad (a_{ij} + b_{jj} n_{ij} + \textstyle\sum_i) \quad &(a'_{kj} + b'_{jj} n_{kj} + \textstyle\sum'_k) \\ &= (a_{kj} + b_{jj} n_{kj} + \textstyle\sum_k)(a'_{ij} + b'_{jj} n_{ij} + \textstyle\sum'_i)\end{aligned}$$

Suppose that $b_{jj} = 0$ and $n_{im} = n_{km} = 0$, for all $m \neq j$ (this is possible without changing the counts N_m, since $s \geq 3$). Then (20) reduces to

$$a_{ij}(a'_{kj} + b'_{jj} n_{kj}) = a_{kj}(a'_{ij} + b'_{jj} n_{ij})$$

since the first probability terms on both sides are assumed to be independent of the counts n_{ij} and n_{kj}, respectively. By first setting $n_{kj} = N_j$ and then setting $n_{ij} = N_j$ we get the following two equations:

$$a_{ij}(a'_{kj} + b'_{jj} N_j) = a_{kj} a'_{ij} \qquad a_{ij} a'_{kj} = a_{kj}(a'_{ij} + b'_{jj} N_j)$$

Subtracting the second from the first equation yields $a_{ij} b'_{jj} N_j = -a_{kj} b'_{jj} N_j$. If $N_j > 0$, then $b'_{jj} = 0$, since $a_{ij}, a_{kj} > 0$. An analogous argument shows that $b'_{jj} = 0$ implies $b_{jj} = 0$.

Consider now the case $j \neq m$. Postulate (7a) states that

$$\begin{aligned}&\mathbb{P}[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j] \\ &\quad = \mathbb{P}[X_{N+1} = k, X_{N+2} = l, X_{N+3} = i | \boldsymbol{X}_N, \boldsymbol{Y}_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]\end{aligned}$$

where we assume that i, l, k are distinct. Again by using postulate (9), it follows that

$$\begin{aligned}&f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_j, \ldots, N_t) \cdot f_{lm}(n_{l1}, \ldots, n_{lt}, N_1, \ldots, N_j + 1, \ldots, N_t) \\ &\quad \cdot f_{kj}(n_{k1}, \ldots, n_{kt}, N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_t) \\ &\quad = f_{kj}(n_{k1}, \ldots, n_{kt}, N_1, \ldots, N_j, \ldots, N_t) \cdot f_{lm}(n_{l1}, \ldots, n_{lt}, N_1, \ldots, N_j + 1, \ldots, N_t) \\ &\quad \cdot f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_t)\end{aligned}$$

Since the second term is the same on each side, this implies

$$(21)\quad (a_{ij}+b_{jj}n_{ij}+\textstyle\sum_i)\;(a''_{kj}+b''_{jj}n_{kj}+\sum''_k) = (a_{kj}+b_{jj}n_{kj}+\sum_k)(a''_{ij}+b''_{jj}n_{ij}+\sum''_i)$$

where now $a''_{ij}=a_{ij}(N_1,\ldots,N_m+1,\ldots,N_j+1,\ldots,N_t)$ and $b''_{jj}=b_{jj}(N_1,\ldots,N_m+1,\ldots,N_j+1,\ldots,N_t)$, etc. The same argument following equation (20) shows that $b''_{jj}=0$ whenever $b_{jj}=0$ and vice versa (assuming that $N_j>0$). From the first part of the proof it follows that $b''_{jj}=b_{jj}(N_1,\ldots,N_m+1,\ldots,N_j,\ldots,N_t)$, and the conclusion of the lemma follows. ■

If the random variables X_n of type j are independent of each other, then $b_{jj}(0,\ldots,N_j,\ldots,0)=0$ for $N_j=1$. It follows from Lemma 2 that, if $b_{jj}(0,\ldots,N_j,\ldots,0)=0$ for $N_j=1$, we have $b_{jj}(N_1,\ldots,N_m,\ldots,N_t)=0$ for all $(N_1,\ldots,N_t)$ with $N_j>0$. Thus, in the case of independence, an observation of an outcome of type j is probabilistically irrelevant for the predictive probability of another outcome of type j, since in this case

$$f_{ij}(0,\ldots,n_{ij},\ldots,0,0,\ldots,N_j,\ldots,0)=a_{ij}(0,\ldots,N_j,\ldots,0)$$

where n_{ij} may be 0 or 1.

On the other hand, if the random variables X_n of a type are not independent, then $b_{jj}(0,\ldots,N_j,\ldots,0)\neq 0$ for $N_j=1$. Lemma 2 implies $b_{jj}(N_1,\ldots,N_m,\ldots,N_t)\cdot b_{jj}(N_1,\ldots,N_m+1,\ldots,N_t)\neq 0$ $(N_j>0)$. This is the main hypothesis of the next lemma.

Lemma 3 . Let $X_1,\ldots,X_{N+1},X_{N+2},X_{N+3}$, $N\geq 1$, be a sequence of outcomes and $Y_1,\ldots,Y_{N+1},Y_{N+2},Y_{N+3}$, $N\geq 1$, a sequence of types such that (6), (7), (8) and (9) hold. Suppose that $N_j>0$, and that for all $1\leq m\leq t$, $b_{jj}(N_1,\ldots,N_m,\ldots,N_t)\cdot b_{jj}(N_1,\ldots,N_m+1,\ldots,N_t)\neq 0$. Then the following three statements are true for any type m:

(1) $b_{jj}(N_1,\ldots,N_m,\ldots,N_t)\cdot b_{jj}(N_1,\ldots,N_m+1,\ldots,N_t)>0$

(2) $\kappa_{ij}(N_1,\ldots,N_m,\ldots,N_t)=\kappa_{ij}(N_1,\ldots,N_m+1,\ldots,N_t)$

(3) *for* $u\neq j$, if $N_u>0$, *then*
$\beta_{ju}(N_1,\ldots,N_m,\ldots,N_t)=\beta_{ju}(N_1,\ldots,N_m+1,\ldots,N_t)$

Proof. *(1)* and *(2)*: Suppose that $m=j$. The relevant predictive probabilities can be given by (18) since $b_{jj}(N_1,\ldots,N_m\ldots,N_t)\cdot$

$b_{jj}(N_1 \ldots, N_m + 1 \ldots, N_t) \neq 0$. Setting $n_{il} = n_{kl} = 0$ for all $l \neq j$ in equation (20) implies

$$(22) \quad \left(\frac{n_{ij}+\kappa_{ij}}{N_j+\sum_j+K_j}\right)\left(\frac{n_{kj}+\kappa'_{kj}}{N_j+1+\sum'_j+K'_j}\right) = \left(\frac{n_{kj}+\kappa_{kj}}{N_j+\sum_j+K_j}\right)\left(\frac{n_{ij}+\kappa'_{ij}}{N_j+1+\sum'_j+K'_j}\right)$$

where $\kappa_{ij} = \kappa_{ij}(N_1 \ldots, N_j, \ldots, N_t)$, $\kappa'_{ij} = \kappa_{ij}(N_1, \ldots, N_j + 1, \ldots, N_t)$, $\beta_{jm} = \beta_{jm}(N_1, \ldots, N_j, \ldots, N_t)$, $\beta'_{jm} = \beta_{jm}(N_1, \ldots, N_j + 1, \ldots, N_t)$, $\sum_j = \sum_{m \neq j} \beta_{jm} N_m$, $\sum'_j = \sum_{m \neq j} \beta'_{jm} N_m$, etc. Equation (22) implies that

$$(n_{ij} + \kappa_{ij})(n_{kj} + \kappa'_{kj}) = (n_{kj} + \kappa_{kj})(n_{ij} + \kappa'_{ij})$$

which simplifies to

$$(23) \quad \kappa_{ij} n_{kj} + \kappa'_{kj} n_{ij} + \kappa_{ij}\kappa'_{kj} = \kappa_{kj} n_{ij} + \kappa'_{ij} n_{kj} + \kappa_{kj}\kappa'_{ij}$$

First setting $n_{kj} = N_j$ and then $n_{ij} = N_j$ yields the following two equations:

$$\kappa_{ij} N_j + \kappa_{ij}\kappa'_{kj} = \kappa'_{ij} N_j + \kappa_{kj}\kappa'_{ij} \quad \kappa'_{kj} N_j + \kappa_{ij}\kappa'_{kj} = \kappa_{kj} N_j + \kappa_{kj}\kappa'_{ij}$$

Subtracting the second equation from the first gives $\kappa_{ij} + \kappa_{kj} = \kappa'_{ij} + \kappa'_{kj}$. Since $s \geq 3$, the same equation has to hold for at least two other pairs of outcomes (i, l), and (k, l) besides (i, k):

$$\kappa_{ij} + \kappa_{kj} = \kappa'_{ij} + \kappa'_{kj}, \quad \kappa_{ij} + \kappa_{lj} = \kappa'_{ij} + \kappa'_{lj}, \quad \kappa_{kj} + \kappa_{lj} = \kappa'_{kj} + \kappa'_{lj}$$

Subtracting the second equation from the first equation yields

$$\kappa_{kj} - \kappa_{lj} = \kappa'_{kj} - \kappa'_{lj}$$

Now adding the third equation implies that $\kappa_{kj} = \kappa'_{kj}$. The same argument applies to all other outcomes. Hence $\kappa_{ij} = \kappa'_{ij}$ for all i and $K_j = K'_j$. Also, since $a_{ij}, a'_{ij} > 0$, we must have $b_{jj}(N_1, \ldots, N_j, \ldots, N_t) \cdot b_{jj}(N_1, \ldots, N_j + 1, \ldots, N_t) > 0$.

Suppose now that $m \neq j$. Then as in (21),

$$(24) \quad \left(\frac{n_{ij} + \sum_{ij} + \kappa_{ij}}{N_j + \sum_j + K_j}\right)\left(\frac{n_{kj} + \sum''_{kj} + \kappa''_{kj}}{N_j + 1 + \sum''_j + K''_j}\right) = \left(\frac{n_{kj} + \sum_{kj} + \kappa_{kj}}{N_j + \sum_j + K_j}\right)\left(\frac{n_{ij} + \sum''_{ij} + \kappa''_{ij}}{N_j + 1 + \sum''_j + K''_j}\right)$$

where $\kappa''_{ij} = \kappa_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_s)$, $\beta''_{ij} = \beta_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_s)$, etc., and $\sum_{ij} = \sum_{r \neq j} \beta_{jr} n_{ir}$ and $\sum''_{ij} = \sum_{r \neq j} \beta''_{jr} n_{ir}$, $\sum_j = \sum_{r \neq j} \beta_{jr} N_r$, $\sum''_j = \beta''_{jm}(N_m + 1) + \sum_{r \neq j, m} \beta''_{jr} N_r$, etc. If $n_{il}, n_{kl} = 0$ for all $l \neq j$, then (24) is the same as (23) with $\kappa''_{ij}, \kappa''_{ik}$ instead of $\kappa'_{ij}, \kappa'_{ik}$. Therefore, by the same argument following (23), $\kappa_{ij} = \kappa''_{ij}$. That argument can also be used to show that $\kappa''_{ij} = \kappa_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_t)$ (just use $N_m + 1$ instead of N_j). It follows that $\kappa_{ij}(N_1, \ldots, N_m, \ldots, N_j, \ldots, N_t) = \kappa_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_t)$, which again implies $b_{jj}(N_1, \ldots, N_m, \ldots, N_j, \ldots, N_t) \cdot b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_t) > 0$.

(3) Let $m = j$. Given the assumptions of the lemma,

$$f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_j, \ldots, N_t) \cdot f_{kj}(n_{k1}, \ldots, n_{kt}, N_1, \ldots, N_j + 1, \ldots, N_t)$$
$$= f_{kj}(n_{k1}, \ldots, n_{kt}, N_1, \ldots, N_j, \ldots, N_t) \cdot f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_j + 1, \ldots, N_t)$$

implies that

$$\left(\frac{\beta_{ju} n_{iu} + \kappa_{ij}}{N_j + \sum_j + K_j}\right)\left(\frac{\beta'_{ju} n_{ku} + \kappa'_{kj}}{N_j + 1 + \sum'_j + K'_j}\right) = \left(\frac{\beta_{ju} n_{ku} + \kappa_{kj}}{N_j + \sum_j + K_j}\right)\left(\frac{\beta'_{ju} n_{iu} + \kappa'_{ij}}{N_j + 1 + \sum'_j + K'_j}\right)$$

provided that $n_{il} = n_{kl} = 0$ for all $l \neq u$. Since, by the first part of the proof, $\kappa_{ij} = \kappa'_{ij}$ for all outcomes i and $K_j = K'_j$, this equation reduces to

$$\kappa_{ij} n_{ku}(\beta'_{ju} - \beta_{ju}) = \kappa_{kj} n_{iu}(\beta'_{ju} - \beta_{ju})$$

If $(\beta'_{ju} - \beta_{ju}) \neq 0$, then $\kappa_{ij} n_{ku} = \kappa_{kj} n_{iu}$, which can only hold for all $0 \leq n_{iu}, n_{ku} \leq N_u$ if $N_u = 0$ (since $\kappa_{iu}, \kappa_{ku} \neq 0$). Hence $\beta_{ju}(N_1, \ldots, N_j, \ldots, N_t) = \beta_{ju}(N_1, \ldots, N_j + 1, \ldots, N_t)$ whenever $N_u > 0$.

Consider now the case $m \neq j$ and recall that $\kappa_{ij} = \kappa''_{ij}$. Equation (24) implies that

$$\left(n_{ij} + \sum_{ij} + \kappa_{ij}\right)\left(n_{kj} + \sum''_{kj} + \kappa_{kj}\right) = \left(n_{kj} + \sum_{kj} + \kappa_{kj}\right)\left(n_{ij} + \sum''_{ij} + \kappa_{ij}\right)$$

Hence, if $n_{ij} = n_{kj} = 0$,

$$(25) \quad \sum_{ij} \sum''_{kj} + \kappa_{kj} \sum_{ij} + \kappa_{ij} \sum''_{kj} = \sum_{kj} \sum''_{ij} + \kappa_{ij} \sum_{kj} + \kappa_{kj} \sum''_{ij}$$

Suppose also that $n_{ir} = n_{kr} = 0, r \neq u$. Then (23) reduces to

$$\beta_{ju} n_{iu} \cdot \beta''_{ju} n_{ku} + \kappa_{kj} \beta_{ju} n_{iu} + \kappa_{ij} \beta''_{ju} n_{ku} = \beta_{ju} n_{ku} \cdot \beta''_{ju} n_{iu} + \kappa_{ij} \beta_{ju} n_{ku} + \kappa_{kj} \beta''_{ju} n_{iu}$$

By the same argument as in the case $m = j$, it follows that $\beta_{ju} = \beta''_{ju}$ provided that $N_u > 0$. The conclusion now follows, since $\beta''_{ju} = \beta_{ju}(N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_t)$ by the same argument as above. ■

Together, the three lemmas imply that the predictive probabilities (10) hold under the assumptions of Theorem 1 whenever $N_j \geq 1$. As explained above, with the help of Lemma 2, Lemma 3 applies if the random variables X_n of a particular type are not independent. This is a hypothesis of the theorem, and thus Lemma 3 together with Lemma 1 yields the representation (10) whenever $N_j \geq 1$. We now show that it also holds if $N_j = 0$. Let $f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t)$ denote the predictive probability for $N_j = 0$.

Lemma 4. Let $X_1, \ldots, X_{N+1}, X_{N+2}, X_{N+3}$, $N \geq 1$, be a sequence of outcomes and $Y_1, \ldots, Y_{N+1}, Y_{N+2}, Y_{N+3}$, $N \geq 1$, a sequence of types such that (6), (7), (8) and (9) hold. Suppose that $\mathrm{N_j} = 0$, and that for all $1 \leq m \leq t$, $b_{jj}(N_1, \ldots, N_m, \ldots, N_t) \cdot b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_t) \neq 0$. Then

$$f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) = \frac{\sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{\sum_{m \neq j} \beta_{jm} N_m + K_j}$$

In particular,

$$f_{ij}(0, \ldots, 0; 0, \ldots, 0) = \frac{\kappa_{ij}}{K_j}$$

Proof. Let $f_{ij}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t)$ denote the predictive probability for $N_j = 1$. Partial exchangeability (7) implies that

$$f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) \cdot f_{kj}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t)$$
$$= f_{kj}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) \cdot f_{ij}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t)$$

It follows that for all $i = 1, \ldots, s$

$$\frac{f_{ij}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t)}{f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t)} = c_j$$

where c_j only depends on j and N_m for all $m \neq j$. Hence,

$$(26)\quad f_{ij}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) = c_j f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t)$$

Let $f_{ij}(0, \ldots, 1, \ldots, 0; N_1, \ldots, 1, \ldots, N_t)$ be the predictive probability for $n_{ij} = 1$ and $N_j = 1$. Then

$$f_{ij}(0, \ldots, 1, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) + \sum\nolimits_{k \neq i} f_{kj}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) = 1$$

Thus, by (26),

$$\begin{aligned} f_{ij}(0, \ldots, 1, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) &= 1 - \sum\nolimits_{k \neq i} f_{kj}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) \\ &= 1 - \sum\nolimits_{k \neq i} c_j f_{kj}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) \\ &= 1 - c_j(1 - f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t)) \end{aligned}$$

Altogether, it follows from Lemma 1 and Lemma 3 that

$$\begin{aligned} f_{ij}(0, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) &= c_j f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) \\ &= \frac{\sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{1 + \sum_{m \neq j} \beta_{jm} N_m + K_j} \end{aligned}$$

and

$$\begin{aligned} f_{ij}(0, \ldots, 1, \ldots, 0; N_1, \ldots, 1, \ldots, N_t) &= 1 - c_j(1 - f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t)) \\ &= \frac{1 + \sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{1 + \sum_{m \neq j} \beta_{jm} N_m + K_j} \end{aligned}$$

Therefore

$$\frac{c_j f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t)}{1 - c_j(1 - f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t))} = \frac{\sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{1 + \sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}$$

This implies

$$(27)\quad f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) \frac{c_j}{1 - c_j} = \sum\nolimits_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}$$

Summing over all $i = 1, \ldots, s$ yields

$$\frac{c_j}{1 - c_j} = \sum\nolimits_{m \neq j} \beta_{jm} N_m + K_j$$

Inserting into (27) gives

$$f_{ij}(0, \ldots, 0; N_1, \ldots, 0, \ldots, N_t) = \frac{\sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{\sum_{m \neq j} \beta_{jm} N_m + K_j}$$

The special case follows if $N_1, \ldots, N_t = 0$. ■

This completes the proof of Theorem 1.

We now turn to the proof of Corollary 1. Suppose that $b_{jj}(0, \ldots, N_j, \ldots, 0) < 0$ for $N_j = 1$. It follows from Lemma 3(1) that $b_{jj}(N_1, \ldots, N_j, \ldots, N_t) < 0$ for all $N_j \geq 1$. But then

$$N_j + \sum\nolimits_{m \neq j} \beta_{jm} N_m + K_j = \frac{1}{b_{jj}} < 0$$

To see that this cannot be the case for arbitrarily large N, just set $N_m = 0$ for $m \neq j$, and let N_j be sufficiently large. Hence $b_{jj}(N_1, \ldots, N_j, \ldots, N_t) > 0$, and thus also $\kappa_{ij} > 0$ for all i. Moreover, $N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j > 0$. Setting $N_m = 0$ for all $m \neq j, l$ and $N_j = 1$, and letting N_l be sufficiently large, shows that $\beta_{jl} \geq 0$. ■

Finally, in order to prove Corollary 2, suppose that (7a) holds if $l = k$. It follows from Theorem 1 that

$$\begin{aligned}
&\left(\frac{n_{ij} + \sum_{r \neq j} \beta_{jr} n_{ir} + \kappa_{ij}}{N_j + \sum_{r \neq j} \beta_{jr} N_r + K_j}\right)\left(\frac{n_{km} + \sum_{r \neq m} \beta_{mr} n_{kr} + \kappa_{km}}{N_m + \beta_{mj}(N_j + 1) + \sum_{r \neq j, m} \beta_{mr} N_r + K_m}\right) \\
&\left(\frac{n_{kj} + \beta_{jm}(n_{km} + 1) + \sum_{r \neq j, m} \beta_{jr} n_{kr} + \kappa_{kj}}{N_j + 1 + \beta_{jm}(N_m + 1) + \sum_{r \neq j, m} \beta_{jr} N_r + K_j}\right) \\
&= \left(\frac{n_{kj} + \sum_{r \neq j} \beta_{jr} n_{kr} + \kappa_{kj}}{N_j + \sum_{r \neq j} \beta_{jr} N_r + K_j}\right)\left(\frac{n_{km} + \beta_{mj}(n_{kj} + 1) + \sum_{r \neq j, m} \beta_{mr} n_{kr} + \kappa_{km}}{N_m + \beta_{mj}(N_j + 1) + \sum_{r \neq j, m} \beta_{mr} N_r + K_m}\right) \\
&\left(\frac{n_{ij} + \sum_{r \neq j} \beta_{jr} n_{ir} + \kappa_{ij}}{N_j + 1 + \beta_{jm}(N_m + 1) + \sum_{r \neq j, m} \beta_{jr} N_r + K_j}\right)
\end{aligned}$$

Set $n_{kr} = 0$ for all $r \neq j, m$. Then the above equation reduces to

$$\beta_{jm}(n_{km} + \beta_{mj} n_{kj} + \kappa_{km}) = \beta_{mj}(n_{kj} + \beta_{jm} n_{km} + \kappa_{kj})$$

Suppose that $n_{kj} = 0$. Setting $n_{km} = 0$ shows that $\beta_{jm}\kappa_{km} = \beta_{mj}\kappa_{kj}$. Thus, if $n_{km} > 0$, $\beta_{jm} n_{km} = \beta_{jm}\beta_{mj} n_{km}$. Hence, if $\beta_{jm} \neq 0$, then $\beta_{mj} = 1$. A similar argument shows that, if n_{km} is set to 0 and n_{kj} is allowed to vary, then $\beta_{mj} \neq 0$ implies $\beta_{jm} = 1$. It follows that either $\beta_{jm} = \beta_{mj} = 0$ or $\beta_{jm} = \beta_{mj} = 1$.

Conversely, if either $\beta_{jm} = \beta_{mj} = 0$ or $\beta_{jm} = \beta_{mj} = 1$, then it is clear from (16) that (7a) holds for $l = k$. ■

References

Achinstein, Peter 1963: 'Variety and Analogy in Confirmation Theory'. *Philosophy of Science*, 30, pp. 207–37.

Aldous, David J. 1985: 'Exchangeability and Related Topics'. In David J. Aldous, Ildar A. Ibragimov, and Jean Jacod, *École d'Été de Probabailités de Saint-Fleur XIII—1983*, pp. 1–198. Edited by P. L. Hennequin. Lecture Notes in Mathematics, 1117. Heidelberg: Springer-Verlag.

Aumann, Robert 1974: 'Subjectivity and Correlation in Randomized Strategies'. *Journal of Mathematical Economics*, 1, pp. 67–96.

Blackwell, David, and James B. MacQueen 1973: 'Ferguson Distributions via Polya Urn Schemes'. *Annals of Statistics*, 1, pp. 353–55.

Carnap, Rudolf 1950: *Logical Foundations of Probability*. Chicago: University of Chicago Press.

—— 1952: *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

—— 1963: 'Variety, Analogy, and Periodicity in Inductive Logic'. *Philosophy of Science*, 30, pp. 222–7.

—— 1971: 'A Basic System of Inductive Logic, Part 1'. In Rudolf Carnap and Richard C. Jeffrey (eds.), *Studies in Inductive Logic and Probability*, vol. I, pp. 33–165. Berkeley and Los Angeles: University of California Press.

—— 1980: 'A Basic System of Inductive Logic, Part 2'. In Jeffrey 1980, pp. 7–155.

Carnap, Rudolf and Wolfgang Stegmüller 1959: *Induktive Logik und Wahrscheinlichkeit*. Vienna: Springer.

Costantini, Domenico 1979: 'The Relevance Quotient'. *Erkenntnis*, 14, pp. 149–57.

—— 1983: 'Analogy by Similarity'. *Erkenntnis*, 20, pp. 103–14.

de Finetti, Bruno 1937: 'La prevision: ses lois logiques ses sources subjectives'. *Annales d l'institut Henri Poincaré*, 7, pp. 1–68. Translated as 'Foresight: Its Logical Laws, Its Subjective Sources' in Henry E. Kyburg, Jr., and Howard E. Smokler (eds.), *Studies in Subjective Probability*, pp. 93–158. Wiley: New York, 1964.

de Finetti, Bruno 1938: 'Sur la condition d'equivalence partielle'. *Actualités Scientifiques et Industrielles*, 739: *Colloques consacré à la théorie des probabilités, VIième partie*, pp 5–18. Translated as

'On the Condition of Partial Exchangeability' in Jeffrey 1980, pp. 193–205.
——1959: 'La probabilita e la statistica nei rapporti con l'induzione, secondo i dwersi punti di vista'. *Corso C.I.M.E. su Induzione e Statistica.* Rome: Cremones. Translated as chapter 9 of his *Probability, Induction and Statistics: The Art of Guessing*, pp. 147–227. New York: Wiley, 1972.
de Finetti, Bruno 1974*: Theory of Probability*, volume 1. London: John Wiley and Sons.
di Maio, Maria Concetta 1995: 'Predictive Probability and Analogy by Similarity in Inductive Logic'. *Erkenntnis*, 43, pp. 369–94.
Diaconis, Persi, and David Freedman 1980: 'De Finetti's Generalizations of Exchangeability'. In Jeffrey 1980, pp. 233–49.
Festa, Roberto 1997: 'Analogy and Exchangeability in Predictive Inferences'. *Erkenntnis*, 45, pp. 89–112.
Good, Irving J. 1965: *The Estimation of Probabilities: An Essay in Modern Bayesian Methods*. Cambridge, MA: MIT Press.
Hesse, Maria 1964: 'Analogy and Confirmation Theory'. *Philosophy of Science*, 31, pp. 319–24.
Hill, Alexandra, and Jeffrey Paris 2013: 'An Analogy Principle in Inductive Logic'. *Annals of Pure and Applied Logic*, 64, pp. 1293–1321.
Jeffrey, Richard C. (ed.) 1980: *Studies in Inductive Logic and Probability*, vol. II. Berkeley and Los Angeles: University of California Press.
Johnson, William E. 1924: *Logic, Part III: The Logical Foundations of Science*. Cambridge: Cambridge University Press.
——1932: 'Probability: The Deductive and Inductive Problems'. *Mind*, 41, pp. 409–23.
Kuipers, Theo A. F. 1984: 'Two Types of Inductive Analogy by Similarity'. *Erkenntnis*, 21, pp. 63–87.
——1978: *Studies in Inductive Probability and Rational Expectation.* Dordrecht: D. Reidel.
Laplace, Pierre Simon 1774: 'Mémoire sur la probabilité des causes par led évènemens'. *Mémoires de Mathématque et Physique, Presentés à l'Académie Royale des Sciences, par divers Savans & lûs dans ses Assemblées, Tome Sixiéme*, 66, pp. 621–56. Translated with commentary by Stephen M. Stigler in Statistical Science, 1 (1986), pp. 359–78.

Link, Godehard 1980: 'Representation Theorems of the de Finetti Type for (Partially) Symmetric Probability Measures'. In Jeffrey 1980, pp. 207–31.

Maher, Patrick 2000: 'Probabilities for Two Properties'. *Erkenntnis*, 52, pp. 63–91.

—— 2001: 'Probabilities for Multiple Properties: The Models of Hesse and Carnap and Kemeny'. *Erkenntnis*, 55, pp. 183–216.

Niiniluoto, Ilkka 1981: 'Analogy and Inductive Logic'. *Erkenntnis*, 16, pp. 1–34.

Romeijn, Jan Willem 2006: 'Analogical Predictions for Explicit Similarity'. *Erkenntnis*, 64, pp. 253–80.

Skyrms, Brian 1993a: 'Analogy by Similarity in Hyper-Carnapian Inductive Logic'. In John Earman, Allen I. Janis, Gerald J. Massey, and Nicholas Rescher (eds.), *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*, pp. 273–83. Pittsburgh: University of Pittsburgh Press.

—— 1993b: 'Carnapian Inductive Logic for a Value Continuum'. *Midwest Studies in Philosophy*, 18, pp. 78–89.

Spohn, Wolfgang 1981: 'Analogy and Inductive Logic: A Note on Niiniluoto'. *Erkenntnis*, 16, pp. 35–52.

Zabell, Sandy L. 1982: 'W. E. Johnson's "Sufficientness" Postulate'. *Annals of Statistics*, 10, pp. 1091–9.

—— 1989: 'The Rule of Succession'. *Erkenntnis*, 31, pp. 283–321.

—— 2011: 'Carnap and the Logic of Inductive Inference'. In Dov M. Gabbay, Stephan Hartmann, and John Woods (eds.), *Handbook of the History of Logic*, pp 265–309. Amsterdam: Elsevier.

