Analogical Predictions for Explicit Similarity

Jan-Willem Romeyn

Abstract

This paper concerns exchangeable analogical predictions based on similarity relations between predicates, and deals with a restricted class of such relations. It describes a system of Carnapian $\lambda\gamma$ rules on underlying predicate families to model the analogical predictions for this restricted class. Instead of the usual axiomatic definition, the system is characterized with a Bayesian model that employs certain statistical hypotheses. Finally the paper argues that the Bayesian model can be generalized to cover cases outside the restricted class of similarity relations.

I Analogy within Carnapian rules

Imagine that the marketing director of a bowling alley is interested in the demographic composition of the crowds visiting her alley. Every evening she records the gender of some of the visitors, and whether they are married or not. Now let us say that on one evening half of the recorded visitors are male and married, and the other half are female and unmarried. Then if a newly arrived visitor is a man, the director may consider it more likely that he will be married than not. But if this visitor is a woman, the director may consider it more likely that she is not married. At least some of the similarity between individuals at the bowling alley is thus explicit in the observation language, namely in their gender, and in making predictions on their marital status this similarity may be employed. In such a case we speak of inductive relevance of explicit similarity in terms of a specific predicate. Such relations are possible because individuals are categorised with predicates from more than one predicate family. The explicit similarity is always with respect to one of these predicate families.

Putting G for gender and M for marital status, the analogical prediction in the example of explicit similarity may be represented in the following way:

$$
\begin{array}{rcl}
G_1^0 & \cap & M_1^1 \\
G_2^1 & \cap & M_2^0 \\
 & \vdots & \\
G_{n-1}^0 & \cap & M_{n-1}^1 \\
G_n^1 & \cap & M_n^0 \\
G_{n+1}^0 & & \\
\hline
M_{n+1}^1 & \text{more likely than} & M_{n+1}^0 \quad .
\end{array}
$$

Here G_i^g with $g = 0$ or $g = 1$ is the record that individual i is male or female, and M_i^m with $m = 0$ or $m = 1$ that this individual is not married or married respectively. The similarity between the individuals with odd index and the further individual $n + 1$ is that all of them satisfy the predicate G^0 from the family G, meaning that they are all male. This similarity is used to derive, from the fact that the odd indexed individuals satisfy the predicate M^1 from the family M, meaning that they are all married, that probably also the individual $n+1$ is married. So the similarity of gender is made explicit in the observation language, and employed for predicting the marital status.

Instead of the two predicate families above, we may imagine that the marketing director categorizes the individuals at the bowling alley according to the division of bachelor, husband, maiden and wife, denoted with the family of predicates Q^q for $q = 0, 1, 2, 3$ respectively. This family is linked to the families G and M according to

$$Q^{2g+m} = G^g \cap M^m, \tag{1}$$

so $Q^0 = G^0 \cap M^0$ is a bachelor, $Q^1 = G^0 \cap M^1$ is a husband, $Q^2 = G^1 \cap M^0$ is a maiden, and $Q^3 = G^1 \cap M^1$ is a wife. Using the single predicate family, the director may derive predictions on gender and marital status from the $\lambda\gamma$ rules of Stegmüller (1973):

$$p(Q_{n+1}^q | E_n) = \frac{n_q + \lambda\gamma_q}{n + \lambda}. \tag{2}$$

Here the expression E_n represents the records of Q-predicates for the first n subjects, and n_q is the number of records of category q within E_n. The parameters γ_q determine the initial expectations over the family Q, and the parameter λ determines the speed with which we change these initial expectations into the recorded relative frequencies of the predicates Q^q. With an assumption of initial symmetry we can fix $\gamma_q = 1/4$ for all q.

With these $\lambda\gamma$ rules concerning Q-predicates we can also derive predictions on the underlying predicate families G and M, using the inverse identifications

$$G^g = Q^{2g} \cup Q^{2g+1}, \tag{3}$$

$$M^m = Q^m \cup Q^{2+m}. \tag{4}$$

With this we can derive the following expressions for predictions on marital status:

$$p(M^1_{n+1}|E_n) = \frac{(n_1+n_3)+\lambda(\gamma_1+\gamma_3)}{n+\lambda}, \tag{5}$$

$$p(M^1_{n+1}|E_n \cap G^0_{n+1}) = \frac{n_1+\lambda\gamma_1}{(n_0+n_1)+\lambda(\gamma_0+\gamma_1)}. \tag{6}$$

The prediction rules thus derived have the same format as the above $\lambda\gamma$ rules. Note that the indices of n refer to the Q-predicates.

On the evening of the example, there are, up to a certain moment, an even number n of visitors at the bowling alley, of which half are husbands and half are maidens:

$$E_n = \bigcap_{i=1}^{n/2}(Q^1_{2i-1} \cap Q^2_{2i}). \tag{7}$$

We therefore have $n_1 = n_2 = {}^n\!/2$ and $n_0 = n_3 = 0$. Then visitor $n{+}1$ parks a car, and upon entering it turns out to be a man. As already suggested in Carnap and Stegmüller (1959: 242-250) and made explicit in Niiniluoto (1981), the symmetric $\lambda\gamma$ rule on family Q predicts a higher probability for this individual being married after incorporating that the visitor is a man than if the gender is unknown:

$$p(M^1_{n+1}|E_n \cap G^0_{n+1}) = \frac{n/2+\lambda/4}{n/2+\lambda/2} > \frac{n/2+\lambda/2}{n+\lambda} = p(M^1_{n+1}|E_n). \tag{8}$$

The $\lambda\gamma$ rule thus shows analogy effects of explicit similarity, in the sense that the similarity of visitor $n+1$ to the present visitors with respect to the family G, the gender, affects the predictions with respect to family M, the marital status.

II Analogical predictions

The above analogy effects are captured in the $\lambda\gamma$ rules, but many more such effects cannot be captured. It may be the case that husbands, Q^1, and bachelors, Q^0, have the disposition to visit the bowling alley together, and that on a particular evening the director has only recorded husbands. Then, apart from the fact that this may make further instances of husbands more probable, we may find an instance of a bachelor more probable than an instance of a maiden or a wife, Q^2 or Q^3, because husbands are more likely to hang out in the bowling alley with their bachelor friends. That is, we consider the presence of husbands more relevant to bachelors than to maidens or wives on the basis of some prior information.

It is easily seen that the $\lambda\gamma$ prediction rules cannot accommodate such differences in relevance among the Q-predicates. For any instance of Q^q, the ratios between the predictions of any two other predicates Q^v and Q^w will not change, because this ratio is given by

$$\frac{p(Q^w_{n+1}|E_n)}{p(Q^v_{n+1}|E_n)} = \frac{n_w + \gamma_w \lambda}{n_v + \gamma_v \lambda}, \tag{9}$$

which is independent of n_q. Therefore analogy effects that hinge on differences in inductive relevance between Q-predicates fall outside the scope of $\lambda\gamma$ rules.

The predictive relevance between Q^v and Q^w may be expressed in terms of an inductive relevance function $\rho(v, w)$. A general expression of analogy by similarity, using the relevance function, is:

$$\rho(q, w) > \rho(q, v) \quad \Rightarrow \quad \frac{p(Q^w_{n+1}|E_{n-1} \cap Q^q_n)}{p(Q^w_n|E_{n-1})} > \frac{p(Q^v_{n+1}|E_{n-1} \cap Q^q_n)}{p(Q^v_n|E_{n-1})}. \tag{10}$$

It must be stressed that this is certainly not the only expression of analogy by similarity, and in particular that the focus differs from that of Carnap (1980: 46-47). The characterization offered here is qualitatively equivalent to Kuipers' characterization in (1984), which is associated with $K_{>G}$ inductive methods in the categorisation of Festa (1997: 232-235). The focus is therefore not on Carnap's and Maher's kind of similarity, which concerns differences between $\rho(v, q)$ and $\rho(w, q)$. On the other hand, I assume in this paper that the function is symmetric:

$$\rho(v, w) = \rho(w, v). \tag{11}$$

Because of this the above expression of relevance is very much akin to that of Carnap and Maher. Note finally that some authors employ a distance function instead of relevances. Strictly speaking this is inadequate, since the relevances need not comply to triangular inequalities.

Many models have been proposed in order to capture analogical predictions based on similarity. The main focus of these models is on an alternative prediction rule concerning Q-predicates that somehow incorporates the relevances. Some of these prediction rules are exchangeable, that is, invariant under permutations of the given Q-predicates, and some are non-exchangeable. Examples of such models are given in Kuipers (1984, 1988), Skyrms (1993), Di Maio (1995) and Festa (1997). However, to my mind analogical predictions are more easily associated with similarity in terms of underlying predicates, here called explicit similarity, than with similarity between predicates directly. Moreover, as it turns out, the use of underlying predicate families is very useful in defining analogical predictions. For these reasons the following employs the underlying predicate families G and M in the construction of the analogical prediction rules for Q.

The models of Carnap, Maher and Niiniluoto do employ underlying predicates. Specifically, Niiniluoto (1981, 1988) uses the structure of underlying predicates to explicate the strengths of the similarity between the Q-predicates. As an example, husbands and bachelors are more similar than husbands and maidens, because the first two have their gender in common, where the second two do not share any underlying predicate. To the extent that this explication of similarity between Q-predicates is adopted in other models, we can say that these other models employ the underlying predicates as well. However, in all these models the relation between the similarities and the prediction rules is rather ad hoc. The predictions of Q are influenced by the similarities, but the explication of the similarity in terms of underlying predicates is itself not used in the construction of the prediction rules. The rules are defined by assigning probabilities to the Q-predicates alone. Probabilities over M and G may be derived from that, but no use is made of the possibility to assign probabilities over the families M and G.

The model of Maher (2000), which is basically an improved version of the model of Carnap and Stegmüller (1959: 251-252), makes more elaborate use of underlying predicates. Maher supposes two predicate families, like G and M, to underly the Q-predicate. He distinguishes two hypotheses on these underlying predicates, one concerning complete independence between them, and one on undifferentiated dependence. He then translates these hypotheses into hypotheses on Q-predicates, and employs the latter in a model of analogical predictions. However, conditional on the hypothesis of statistical dependence, the model of Maher comes down to a single $\lambda\gamma$ rule for the Q-predicates. By contrast, the present model employs predictions on underlying predicates in the case of dependence as well. Moreover, and perhaps more importantly, the present model elucidates the exact relation between inductive relevances ρ and the statistical dependencies between underlying predicates. Specifically, the relevance relations ρ, which are assumed at the start of the update, are related to the parameters of the model. On this point the model differs from Maher's (2000) model and the other models discussed in Maher (2001), which are not directly related to assumed relevance relations ρ.

The model of this paper thereby is restricted in a certain way. It provides analogical predictions that cannot be captured by the single $\lambda\gamma$ rule, but it considers only a limited set of relevance relations. As an example, consider the husbands and bachelors who like to go bowling together, and prefer not to have female company. In terms of relevance functions,

$$\rho(0,1) = \rho(1,0) > \rho(0,2) = \rho(1,2) = \rho(0,3) = \rho(1,3). \tag{12}$$

That is, the relevances of husbands and bachelors to each other are equal, and larger than relevances between individuals of different gender. Let us further say that if wives visit the bowling alley, they are likely to bring their husbands, who then also invite their bachelor friends. But the wives typically do not invite their maiden friends. Similarly, if maidens visit the bowling alley, they are likely to be together with the bachelors, who in turn bring along some husbands, but the maidens do not usually invite any wives. That is,

$$\rho(2,3) = \rho(3,2) < \rho(2,0) = \rho(2,1) = \rho(3,0) = \rho(3,1), \tag{13}$$

or in words, the relevances of wives and maidens to each other are equal, and smaller than relevances between individuals of different gender. Note that due to the symmetry of the relevance function, the four equal relevances in expressions (12) and (13) are the same.

As said, this example is one in a set of similar cases. The common element is that the relevances between categories with different gender are all equal, and that the relevances between categories within the genders may vary. Defining

$$\forall m, m' \in \{0,1\}: \qquad \rho_G = \rho(m, 2+m'), \tag{14}$$

$$\forall g \in \{0,1\}: \qquad \rho_M^g = \rho(2g, 2g+1), \tag{15}$$

the similarity relations are in effect characterized by three relevances, ρ_G, ρ_M^0 and ρ_M^1, representing the relevances between individuals of different gender, the relevance between bachelors and husbands, and the relevance between maidens and wives respectively. These relevances may have any ordering in size. The subclass of cases thus defined, for which the relevance relations between categories of different gender do not differ, are exactly the cases of analogy by similarity that can be made explicit in terms of gender. In the following I therefore refer to them as cases of explicit similarity.

Summing up, the aim of this paper is to provide a rule for analogical predictions based on explicit similarity with respect to a specific predicate, namely G, to connect the relevance relations ρ_G, ρ_M^0 and ρ_M^1 to parameters in this prediction rule, and finally to give a proper statistical underpinning for it.

III A model for explicit similarity

This section presents a system of $\lambda\gamma$ rules that models the intended analogical predictions. It is shown that the system generalizes the analogy effects that are captured in single $\lambda\gamma$ rules. The function of the parameters in the system is explained, and a numerical example is provided.

The system of prediction rules offers separate entries for instances of the family M for individuals satisfying either of the two predicates of the family G. This is expressed in the following:

$$\begin{aligned} p(G^g_{n+1}|E_n) &= \frac{n_{Gg} + \lambda_G \gamma_{Gg}}{n_G + \lambda_G}, & (16)\\ p(M^m_{n+1}|E_n \cap G^g_{n+1}) &= \frac{n^g_{Mm} + \lambda^g_M \gamma^g_{Mm}}{n^g_M + \lambda^g_M}. & (17) \end{aligned}$$

The indexed numbers n can all be derived from E_n using the translations (3) and (4). In particular, we have the total number of records on gender $n_G = n$, the number of records of males and females, $n_{Gg} = n_{2g} + n_{2g+1}$ for $g = 0, 1$, which is the same as the number of records on marital status given a certain gender, $n^g_M = n_{Gg}$, and finally the number of records for a specific gender and marital status, $n^g_{Mm} = n_{2g+m}$ for $g, m \in \{0, 1\}$.

The above system consists of three prediction rules, one that concerns individual $n + 1$ in the family G, and two that concern the family M, conditional on the individual satisfying G^0 and G^1 respectively. With these predictions we can construct a prediction rule for Q-predicates:

$$\begin{aligned} p(Q^q_{n+1}|E_n) &= p(G^g_{n+1}|E_n) \times p(M^m_{n+1}|E_n \cap G^g_{n+1}) \\ &= \frac{n_{Gg} + \lambda_G \gamma_{Gg}}{n_G + \lambda_G} \times \frac{n^g_{Mm} + \lambda^g_M \gamma^g_{Mm}}{n^g_M + \lambda^g_M}. & (18) \end{aligned}$$

As will be seen below, these predictions for Q-predicates can capture explicit similarity. But note first that the above system is a generalization of the single $\lambda\gamma$ rule for Q-predicates. By writing the numbers n in terms of the n_q, by identifying

$$\begin{aligned} \lambda_G &= \lambda, & (19)\\ \gamma_{2g+m} &= \gamma_{Gg}\gamma^g_{Mm}, & (20) \end{aligned}$$

and finally by choosing

$$\lambda^g_M = \lambda_G \gamma_{Gg}, \tag{21}$$

the system of rules generates the very same predictions that are generated by the single $\lambda\gamma$ rule of equation (2). Note also that on the level of Q-predicates the predictions are exchangeable, whatever the values of the parameters.

Analogical predictions for explicit analogy can be obtained by choosing the values of the parameters λ^g_M different from those in equation (21). To explain this, let me first reformulate explicit similarity in terms of probabilities over the underlying predicate families G and M. First, the higher relevance between husbands and bachelors means that the effect of updating with the male gender of an individual must be larger than the effect of updating with the marital status conditional on the individual being male. In case we update for a husband, for example, the probability for further instances of males may strongly benefit from the husband being male, while the husband's being married does not make bachelors much less likely. In similar fashion, the lower relevance between wives and maidens means that the effect of updating with the female gender of, say, a maiden is much smaller than the effect of updating with the marital status of the maiden conditional on her being female. On finding a maiden, the profit that the wives derive from the fact that the maiden is female is then overcompensated by the loss that stems from the fact that contrary to the wives, the maidens are not married.

In the $\lambda\gamma$ rules of Carnap, the reluctance to adapt probabilities to new observations is reflected in the size of λ. In the above formulation, it is exactly differences in the reluctance to adapt probability assignments that leads to analogy effects. The above paragraph therefore suggests that we can connect the differences in relevance with specific differences between the values of the parameters λ_G and λ^g_M. As it turns out, we can identify a correspondence between parameter inequalities and inequalities of relevance functions. Normalizing the size of the relevances for the number of Q-predicates N, so that in this case $N = 4$, these correspondences can be translated into rather simple relations:

$$\lambda_G = \rho_G N, \tag{22}$$

$$\lambda^g_M = \rho^g_M \gamma_{Gg} N. \tag{23}$$

In updating with observations on the family G, we may be more or less prepared to adapt our expectations concerning gender, which is reflected in a low or high value for λ_G respectively. Similarly, conditional on the observation of G^g, we may be more or less prepared to adapt our expectations on an observation concerning M, which is reflected in the value of λ^g_M. With these variations in the willingness to adapt probabilities, we can model explicit analogy of gender.

Let me make explicit the relation between the values of the λs and the relevances between predicates of equal gender for the specific case of husbands and bachelors. Recall that we have chosen $\rho_G < \rho^0_M$. In the model this relevance relation is supposed to be assumed at the start of the update. Now to encode this relevance relation in the system of prediction rules, we must according to the above equations choose $\gamma_{G0}\lambda_G < \lambda^0_M$. With these parameter values, the observation of a male strongly enhances the probability for further males, while the prediction for marital status conditional on males grows much slower with the observation of the males having a specific marital status. In the resulting predictions with respect to the predicates Q, this has the combined effect that observations of husbands and bachelors are mutually beneficial. This is because the observation of the common element of these predicates, their gender, affects the expectations much more than the observation that distinguishes the one from the other, their marital status.

Let us say that one night at the bowling alley the first three visitors are husbands, after which three maidens enter:

$$E_n = Q_1^1 \cap Q_2^1 \cap Q_3^1 \cap Q_4^2 \cap Q_5^2 \cap Q_6^2. \tag{24}$$

For the example of equations (12) and (13) we have $\rho_M^0 > \rho_G > \rho_M^1$. In particular, we may choose $\rho_M^0 = 4$, $\rho_G = 1$ and $\rho_M^1 = 1/2$, and accordingly fix the following values for the parameters in the system of prediction rules:

$$\begin{aligned}
\lambda_G &= 4, \\
\lambda_M^0 &= 8, \\
\lambda_M^1 &= 1, \\
\gamma_{G0} &= 1/2, \\
\gamma_{M0}^0 &= \gamma_{M0}^1 = 1/2.
\end{aligned}$$

The predictions that can be generated with the above parameter values then show the analogical effects that can be expected on the basis of the corresponding values of the relevance function:

Number n	0	1	2	3	4	5	6
Observations q	-	1	1	1	2	2	2
$p(Q_{n+1}^0 \mid E_n)$	0.25	0.27	0.27	0.26	0.23	0.20	0.18
$p(Q_{n+1}^1 \mid E_n)$	0.25	0.33	0.40	0.45	0.40	0.35	0.32
$p(Q_{n+1}^2 \mid E_n)$	0.25	0.20	0.17	0.14	0.28	0.37	0.44
$p(Q_{n+1}^3 \mid E_n)$	0.25	0.20	0.17	0.14	0.09	0.07	0.06

As can be seen from these predictions, the husbands are positively relevant to the bachelors, while the maidens are negatively relevant to the wives. As is to be expected from the fact that the predictions are exchangeable, this effect wears off as the number of records increases, but it only reaches zero at infinity.

Let me stress once again an important aspect to the model of analogical predictions defined in this section, namely that inductive relevances serve as explicit input to the prediction rules. In this sense the model is similar to the models of Niiniluoto (1981) and Kuipers (1984), while it differs from the models of Festa (1996) and Maher (2000). In these latter models, there is no direct access, in terms of input parameters, to the inductive relevances that may be assumed. However, whether this aspect of the accessibility of inductive relevances can be considered an advantage depends on the perspective we take on inductive logic.

IV Problematic aspects

This section discusses the fact that the system shows two distinct asymmetries in dealing with the families G and M, and motivates a difference in methodology that sets this treatment apart from the Carnapian tradition.

First I discuss whether the above system of rules preselects an order in the underlying predicate families. Note that in the above system of rules, we can only directly adapt the predictions for the marital status of individuals if we already know their gender. For example, if we know that only unmarried people drive sport scars, and we see a visitor arriving at the bowling alley in such a car before having determined her or his gender, it is not immediately clear how we must adapt the prediction rules. Accordingly, we cannot directly use the system to predict the gender of the visitors on the basis of their being unmarried.

All this is not to say that the system of prediction rules breaks down if the order of the observations is reversed. The system does assign a probability to the gender of a visitor conditional on this visitor having a certain marital status, and it also assigns a probability to the marital status of a visitor unconditionally. Both can be derived from the prediction rules (16) and (17) together with the law of total probability. It is just that the calculations become rather intricate if we update on marital status first, because adapting the system of rules to records of marital status independently of gender is a messy operation. Moreover, as it turns out the numerical values of the predictions do not change under permutations in the order of the underlying predicates. That is to say, the system is still exchangeable with respect to the underlying predicates. Unfortunately, an argument for that can only be given in section VI. For now the main thing is that the system of rules does not necessitate a specific order in the observations to obtain numerical values for the predictions.

Leaving the matter of order and order dependence aside for the moment, it may be noted that there is yet another way in which the above systems treat the underlying predicates differently. When it comes to expressing relevance relations, there is an undeniable asymmetry in the predicates G and M: the systems are perfectly suitable for determining the relevance relation of some

Q-predicate to the other Q-predicate with the same gender, ρ^g_M, relative to the relevance relation of this Q-predicate to the two Q-predicates of the opposite gender, ρ_G. But, swapping the predicate classes of G and M themselves, the system is not at all suitable to determine the relevance relations ρ^m_G relative to the relations ρ_M. In short, the system models explicit similarity of gender, and not of marital status. In setting up the system, we must choose which of these two complexes of relevance relations will be allowed implementation. Therefore, while the system of rules is exchangeable in the sense of invariant under permutations of the order of the observations, it is certainly not suitable for expressing analogy effects after swapping the predicates.

Now in some cases, a natural priority is suggested by the underlying predicate families themselves. One of the two is sometimes more directly observable than the other, or epistemically prior in some other way. For the purpose of adapting the system, and for modelling explicit similarity, such considerations may guide our choice. But there remains an inherent asymmetry in the expressible relevance relations, and in this sense the present model of analogical predictions is weaker than, for example, the model of Maher (2000). It is hoped that this disadvantage is compensated by the new perspective that the model offers, and the new possibilities that may result from that. As section VII suggests, the asymmetry between predicates can eventually be overcome.

Let me turn to the second problematic aspect of the above system, which is that so far, it lacks an axiomatic underpinning. In Carnapian inductive logic the aim is to derive, from a chosen language or algebra and the notion of logical probability, a class of probability assignments over the algebra that describes all rationally permissible predictions. But the above system of rules has been introduced without any such derivation, and in this sense seems entirely ad hoc. It is not even clear whether the probability assignment over the algebra that is entailed by the above system is internally consistent. The remainder of this paper is aimed at solving this problem. To provide the further underpinning of the system of prediction rules, and to prove its internal consistency, the next two sections specify a class of Bayesian statistical models that underlie the proposed systems of rules. In the remainder of this section, I discuss this modelling perspective, and its relation to more traditional Carnapian methods.

As used here, the Bayesian statistical models make use of hypotheses to define the prior probability assignment over the algebra, and they employ Bayesian updating to incorporate observations into this assignment. The system of prediction rules is thus not based on the algebra or language we have chosen, or on further principles or predictive properties we may assume. Instead, we define the inductive prediction rule by partitioning the algebra into a specific set of statistical hypotheses, and by stipulating a class of prior probability assignments over them. At the start we can choose a specific configuration of relevance relations, which may be encoded in a prior probability assignment. This signals an important methodological difference between the present paper and most papers from the Carnapian tradition. The present paper maintains that hypotheses and priors can be chosen freely, and that there are no restrictions implicit in the statistical framework. Relatedly, in this paper there is no attempt to provide a rationalisation for the choice of hypotheses or the prior probability. The hypotheses and prior are taken to exhibit the inductive assumptions underlying the analogical predictions, much like premises in a deductive inference. Attempting to justify hypotheses and prior falls outside the reconstruction of analogical predictions as a statistically sound, or logically valid method.

Adopting this perspective on analogical predictions may look like a cheap escape from a challenging problem. Surely it is much harder to give a set of axioms that have an intuitive appeal or some independent justification, from which the exact class of all rationally permissible analogical predictions can be derived. While searching for these axioms and rationalisations is a worthwhile and venerable task, I side with the criticisms towards such axiomatic methods for analogical predictions, as can be found in Spohn (1981) and Niiniluoto (1988): it may be too ambitious to aim for the definitive class of all rational probability assignments that capture analogical considerations. It is more in line with an emphasis on local inductive practice, as recently discussed in Norton (2003), to propose a collection of models only, and to decide about the exact nature of analogical predictions on a case by case basis. This perspective resembles that of Bovens and Hartmann (2003), who advocate a kind of philosophical engineering as opposed to a quest for first principles.

There are some important advantages to providing statistical models that underly the analogical predictions. First of all, the models connect research in analogical prediction rules with Bayesian statistical inference. I think it is important to bring these research traditions closer together. Second, as will be seen below, extending the models to more predicate families, or to more predicates within given families, is a straightforward operation in the statistical model. It thus turns out that these models are very easy to generalise. Third, the statistical models clarify that the system of prediction rules is really invariant under permutations of the order of the underlying observations. In other words, the statistical underpinning settles the issue of the exchangeability of the underlying observations.

Finally, and perhaps most importantly, the statistical models suggest a more general model of analogical predictions, which accommodates analogical predictions based on more general relevance relations than the ones considered above. We may for example consider bachelors more relevant to maidens than to wives, and it turns out that statistical models offer a rather natural place for relevance relations of this kind. Eventually the use of the statistical model leads to a model of analogical predictions based on a completely general relevance function. This development, however, cannot be dealt with in the present paper.

V Statistical underpinning for $\lambda\gamma$ rules

This section discusses an observation algebra for Q-predicates, and the statistical underpinning of the $\lambda\gamma$ rule for these predicates. The discussion prepares for the statistical underpinning of the system of $\lambda\gamma$ rules in the next section, which employs the basic partition of this section in threefold.

Let me first introduce a representation of records of Q-predicates in terms of a so-called observational algebra. Let K be the set of possible values for q, and let K^∞ be the space of all infinite sequences e of such values:

$$e = q_1 q_2 q_3 \ldots \tag{25}$$

The observation algebra, denoted $\mathcal{Q}$, consists of all possible subsets of the space K^∞. If we denote the i-th element in the sequences e and e_n with $e(i)$ and $e_n s(i)$ respectively, we can define an observation Q_i^q as an element of the algebra $\mathcal{Q}$ as follows,

$$Q_i^q = \{e \in K^\infty : \ e(i) = q\}, \tag{26}$$

and a finite sequence of observations $E_n^{e_n}$ as follows,

$$E_n^{e_n} = \bigcap_{i=1}^{n} Q_i^{e_n(i)}. \tag{27}$$

Records of visitors at the bowling alley refer to such subsets. Note that there is a distinction between the observations, which are elements of the algebra $\mathcal{Q}$, and the values of the observations, which are natural numbers.

Statistical hypotheses can also be seen as elements of the algebra. If we say of a statistical hypothesis h that its truth can be determined as a function $W_h(e)$ of an infinitely long sequence of observations e, writing $W_h(e) = 1$ if h is true for the sequence e and $W_h(e) = 0$ otherwise, then we can define hypotheses as subsets of K^∞:

$$H = \{e \in K^\infty : \ W_h(e) = 1\}. \tag{28}$$

A partition is a collection of hypotheses, $\mathcal{D} = \{H_\theta\}_{\theta \in D}$, defined by the following condition for the indicator functions W_{h_θ}:

$$\forall e \in K^\infty \; \exists! \theta : \quad W_{h_\theta}(e) = 1. \tag{29}$$

The hypotheses here are point hypotheses. The condition entails that the hypotheses H_θ are mutually exclusive and jointly exhaustive sets in K^∞, parameterized by a point vector θ in an as yet unspecified space D.

Probability assignments are defined for all the elements of the observational algebra $\mathcal{Q}$. The probability assignment can be adapted to a sequence of observations E_n by conditioning the original probability assignment p on these observations:

$$p(\cdot) \quad \rightarrow \quad p(\cdot|E_n). \tag{30}$$

Both the probabilities assigned to observations, and the probabilities assigned to hypotheses can be adapted to new observations in this way.

The schemes of this paper employ observational hypotheses for generating the predictions $p(Q^q_{n+1}|E_n)$. To calculate the predictions, we may employ a partition of hypotheses and the law of total probability:

$$p(Q^q_{n+1}|E_n) = \int_D p(H_\theta|E_n)\, p(Q^q_{n+1}|H_\theta \cap E_n)\, d\theta. \tag{31}$$

The probability function over the hypotheses is a so-called posterior probability, $p(H_\theta|E_n)d\theta$. This probability is obtained by conditioning a prior probability $p(H_\theta)d\theta$ on the observations E_n. The terms $p(Q^q_{n+1}|H_\theta \cap E_n)$ are called the posterior likelihoods of the hypotheses H_θ, which are defined for observations Q^q_{n+1}. The prediction is obtained by weighing these posterior likelihoods with the posterior density over the hypotheses.

To characterize the partition that renders exchangeable predictions, define the relative frequency of the observation results $q \in K$ in a sequence e:

$$f_q(e) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} W_q(e(i)), \tag{32}$$

in which $W_q(e(i)) = 1$ if $e(i) = q$, and $W_q(e(i)) = 0$ otherwise. Taking θ to be a real-valued vector, we can define W_{h_θ} as follows:

$$W_{h_\theta}(e) = \begin{cases} 1 & \text{if } \forall q \in K : f_q(e) = \theta_q, \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

The hypotheses H_θ form a so-called simplex, associated with a hypersurface $C = \{\theta \in [0,1]^{|K|} | \sum_q \theta_q = 1\}$ in a $|K|$-dimensional space. For $|K| = 4$, this hypersurface is a tetrahedron. We can further define $W_{h_{\neg\theta}}(e) = 1$ if $f_q(e)$ is undefined for any of the $q \in K$, and $W_{h_{\neg\theta}}(e) = 0$ otherwise. The collection of hypotheses $\mathcal{C} = \{H_{\neg\theta}, \{H_\theta\}_{\theta \in C}\}$ is a partition of hypotheses concerning the relative frequencies of $q \in K$.

We can now provide the likelihoods associated with the partition that renders exchangeable predictions. First we assume that $p(H_{\neg\theta}) = 0$, which states that the observations have some convergent relative frequency. The likelihoods of $H_{\neg\theta}$ may then be left unspecified. The likelihoods of H_θ may be defined by taking the long run relative frequencies θ_q as chances on predicates Q^q at every single observation:

$$\forall n \geq 0 : \quad p(Q^q_{n+1} | H_\theta \cap E_n) = \theta_q. \tag{34}$$

So the likelihoods do not depend on the observations E_n. The prior probability over the hypotheses H_θ can be chosen freely. According to De Finetti's representation theorem, there is a one-to-one mapping between exchangeable prediction rules and prior probability densities over partition $\mathcal{C}$ with these likelihoods.

Because the single $\lambda\gamma$ prediction rules are exchangeable, they can also be characterized by a specific class of densities over $\mathcal{C}$. This turns out to be the class of so-called Dirichlet densities:

$$p(H_\theta) \quad \sim \quad \prod_q \theta_q^{(\lambda\gamma_q - 1)}. \tag{35}$$

On assuming this prior, the resulting predictions are the $\lambda\gamma$ prediction rules with the corresponding parameter values. See Festa (1993: 57-71) for further details. So the $\lambda\gamma$ rules can be generated if we assume a partition of hypotheses $\mathcal{C}$ and its associated likelihoods $p(Q^q_{n+1} | H_\theta \cap E_n) = \theta_q$, and a prior probability density $p(H_\theta)$ from the Dirichlet class.

We can now reformulate the aims of this paper. To capture exchangeable predictions based on analogy by similarity of gender, we must effectively define a prior density over partition $\mathcal{C}$ that is not Dirichlet, and that somehow incorporates analogical effects. Intuitively, we need a prior over the tetrahedron C that has an internal twist: within the triangular segments with hypotheses that have relatively high likelihoods for Q^1, we must allocate more prior probability to those hypotheses that also have relatively high likelihoods for Q^0, and similarly, within the triangular segments with hypotheses that have relatively low likelihoods for Q^1, we must allocate more prior probability to those hypotheses that have relatively low likelihoods for Q^0. With such a twisted prior density, we effectively favour the probability of Q^0 over those of Q^2 and Q^3 whenever we update with Q^1.

On the level of Q-predicates, the system of $\lambda\gamma$ rules defined in equations (16) and (17) is exchangeable, just like the single $\lambda\gamma$ rule. It can therefore be represented as a class of prior densities over $\mathcal{C}$. To find the statistical models underlying the system of rules, we must thus find the exact class of prior densities over $\mathcal{C}$ with which these systems can be represented. However, this class of priors is very hard to define if we only have recourse to the parameter components in the space C. Even if we knew what function satisfies the features sketched above, it is not easy to formulate this prior in such a way that we can actually derive the system of rules. For this reason it is worthwhile to look for an alternative framework. The following proposes a transformation of the partition $\mathcal{C}$ into the partition $\mathcal{A}$. This latter partition comprises exactly the same hypotheses, but casts these in a different parameter space. Within that space we can derive the analogical predictions of section III.

VI Analogy partition

This section proposes a transformation of the algebra $\mathcal{Q}$ into one on observations of predicates from the underlying families G and M. After that the hypotheses and densities that result in the system of $\lambda\gamma$ rules can be specified.

First we must define a space on which the algebra for records concerning G and M can be defined. Taking L as the set of ordered pairs $\langle g, m \rangle$, we can define the space L^{∞} of all infinitely long ordered sequences u of such observations:

$$u = g_1 m_1 \, g_2 m_2 \, g_3 m_3 \, \ldots \tag{36}$$

The record that the individual i is a husband, $q_i = 1$, can then be written as two consecutive records in a sequence u, namely $g_i = 0$ and $m_i = 1$, meaning that the individual i is recorded to be male and married. More generally, we can identify all infinite strings of observations $e \in K^{\infty}$ with some infinite string $u \in L^{\infty}$. Using $u(t)$ as the t-th number in the sequence u, we can construct

$$\begin{aligned} e(i) &= 2g_i + m_i, \\ u(2i-1) &= g_i, \\ u(2i) &= m_i. \end{aligned}$$

In this way every sequence e is mapped onto a unique sequence u, and every such u can be traced back to the original e.

We can now define the algebra $\mathcal{R}$ for records concerning the predicate families G and M on the basis of the space L^{∞}. The following elements generate this algebra:

$$G_i^g = \{u \in L^{\infty} : \, u(2i-1) = g\}, \tag{37}$$

$$M_i^m = \{u \in L^{\infty} : \, u(2i) = m\}. \tag{38}$$

The sets $G_i^g \cap M_i^m$ thus contain all those infinitely long sequences u that have the number g and m in the positions $2i-1$ and $2i$. The relations between the families Q, G and M are therefore as specified in equation (1). For future reference, sequences of records in $\mathcal{R}$ that correspond to a specific e_n are here denoted $S_n^{e_n}$.

The idea of the hypotheses concerning the underlying predicate families is essentially the same as for those concerning Q-predicates. We may again partition the above observational algebra into hypotheses concerning relative frequencies. However, the relative frequencies in the family M must in this case be related to the result in the family G. We may define the following relative frequencies:

$$f_g(u) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} W_g(u(2i-1)), \tag{39}$$

$$f_m^g(u) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} W_g(u(2i-1))W_m(u(2i))}{\sum_{i=1}^{n} W_g(u(2i-1))}. \tag{40}$$

Here $W_r(u(t)) = 1$ if $u(t) = r$ and $W_r(u(t)) = 0$ otherwise. The function f_g simply gives the relative frequency of results g within the observations with respect to G in the sequence u. But the function $f_m^g(u)$ is somewhat more complicated. It returns, for every u, the relative frequency of results m for observations with respect to the family M, conditional on the observed individual belonging to the category g within the family G. This is the relative frequency of m conditional on g within u, or the conditional relative frequency for short.

We are now in a position to define the analogy partition $\mathcal{A}$ for predictions concerning the predicate families G and M. The hypotheses in this partition employ the conditional relative frequencies in order to pick up the exact statistical dependency between the two families. Let α_g and α_{gm} be the parameters labelling these hypotheses, and define

$$W_{h_\alpha}(u) = \begin{cases} 1 & \text{if } f_g(u) = \alpha_g \text{ and } f_m^g = \alpha_{gm} \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

and then define the hypotheses $H_\alpha = \{u \in L^\infty | W_{h_\alpha}(u) = 1\}$. Again define $H_{\neg\alpha}$ as the set of all u for which one of the three relative frequencies in equations (39) or (40) does not exist. The analogy partition is then given by $\mathcal{A} = \{H_{\neg\alpha}\} \cup \{H_\alpha\}_{\alpha \in A}$. Here the parameter $\alpha = \langle \alpha_0, \alpha_1, \alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11} \rangle$ lies in the set $A = \{\alpha \in [0,1]^6 \mid \alpha_0 = 1 - \alpha_1, \alpha_{00} = 1 - \alpha_{01}, \alpha_{10} = 1 - \alpha_{11}\}$.

The likelihoods of the hypotheses on the underlying predicates are given by these relative frequencies and conditional relative frequencies:

$$p(G^g_{i+1}|H_\alpha \cap S^e_i) = \alpha_g, \tag{42}$$

$$p(M^m_{i+1}|H_\alpha \cap S^e_i \cap G^g_{i+1}) = \alpha_{gm}. \tag{43}$$

It may be noted that observations in the sequence S^e_i do not influence the likelihoods of H_α, but that G^g determines which of the α_{gm} must be used as the likelihood for M^m. In this sense, the likelihoods for the family M depend on earlier observations in the family G. Note also that we can write

$$\begin{aligned} p(Q^{2g+m}_{i+1}|H_\alpha \cap S^e_i) &= p(G^g_{i+1} \cap M^m_{i+1}|H_\alpha \cap S^e_i) \\ &= \alpha_g\, \alpha_{gm}. \end{aligned} \tag{44}$$

The likelihoods for the separate families G and M therefore also imply likelihoods for the family Q, and with that also unconditional likelihoods for the family M.

It is useful to consider the parameter space A for the above partition in some more detail, and relate it to the parameter space C. First recall that pairs of the parameter components of α sum to one. The parameter space A is therefore built up from a separate simplex C_G for the two parameters α_g, and two simplexes C_{gM} for the four parameters α_{gm}. We can write

$$A = C_G \times C_{0M} \times C_{1M}. \tag{45}$$

Like the original simplex C for $|K| = 4$, the parameter space A therefore has three independent components. In fact, following the above expression for the likelihoods, the parameter space C can be obtained from A by a simple transformation:

$$\theta_{2g+m} = \alpha_g \alpha_{gm}. \tag{46}$$

When it comes to the statistical hypotheses, the partitions $\mathcal{A}$ and $\mathcal{C}$ are thus equivalent. However, they employ different parameter spaces, and therefore provide access to different classes of prior probability functions. Specifically, the space C is a tetrahedron, which is transformed into a unit cube A. This is the transformation intended at the end of section V.

The prior probability assignment over $\mathcal{A}$ that generates the system of $\lambda\gamma$ rules can now be made precise. It is noteworthy that the partition $\mathcal{A}$ consists of three separate and orthogonal dimensions. However, it is not yet clear whether these three dimensions can be treated independently, let alone that they result in such prediction rules. To establish the independence, it must be determined under what conditions the updates and predictions over the parts do not affect each other. As it turns out, independence is guaranteed if we assume that the prior probability density is factorisable:

$$p(H_\alpha) = p_G(\alpha_g)\, p_{0M}(\alpha_{0m})\, p_{1M}(\alpha_{1m}). \tag{47}$$

In that case updating in one of the dimensions leaves the functions in the other two dimensions unchanged. The separate dimensions in A may then be associated with the separate $\lambda\gamma$ rules of 16 and 17.

In order to derive these $\lambda\gamma$ rules, we must assume more than factorisability. We must assume that the prior densities over the separate dimensions are members of the Dirichlet class:

$$p(H_\alpha) \sim \prod_g \alpha_g^{(\lambda\gamma_g - 1)} \prod_m \alpha_{0m}^{(\lambda_M^0 \gamma_{Mm}^0 - 1)} \prod_m \alpha_{1m}^{(\lambda_M^1 \gamma_{Mm}^1 - 1)}. \tag{48}$$

From here onwards the derivation of the separate $\lambda\gamma$ prediction rules runs entirely parallel to the derivation of a single rule. Again, the details for this derivation may be found in Festa (1993: 57-71).

The statistical model shows that the predictions resulting from them are indeed exchangeable, meaning that the predictions are invariant under permutations of the order of observations. Since we can assign a likelihood for an observation M_i^m before G_i^g on every hypothesis H_α, we can adapt the probability over $\mathcal{A}$ for these unconditional observations of M_i^m in the same way as that we can adapt the probability upon observing G_i^g. Both updates are simply multiplications with the likelihood functions. In the Bayesian model, there is therefore no principled restriction on the order of the observations, and in this sense the Bayesian model offers a wider framework than the system of $\lambda\gamma$ rules. The restriction only shows up as the fact that the integrals for the predictions based on the Bayesian model cannot be solved analytically, in the form of a system of prediction rules, if the observations M_i^m occur before G_i^g.

Let me return to the relation between the parameter spaces C and A. Recall that the class of Dirichlet densities over C corresponds to a special class of Dirichlet densities over A, which results in the predictions determined by equation (21). This follows from the fact that for this choice of parameters, the system of rules boils down to a single $\lambda\gamma$ rule. At the level of the partitions, however, we may also transform the Dirichlet density over C by means of the relations (46), and multiply the transformed density with the Jacobian determinant of the transformation matrix. This results in the very same equivalence. On the other hand, there are many more Dirichlet densities over A that cannot be captured by the Dirichlet densities over C in this way. Transforming these densities over A back to C is a less clear-cut operation, and the resulting densities over C do not fall within any special class of densities. The transformation of C into A has in this way provided access to a new class of prior densities.

As suggested, the above perspective opens up the possibility of modelling many other analogical predictions. We may consider densities over A that are not Dirichlet, and more specifically, that are not factorisable. However, these latter analogical predictions can only be dealt with in a future paper. The next section only presents a brief sketch of these possibilities. For present purposes, the main point is that the system of rules has been connected to a range of statistical models: the existence of these models ensures the consistency of the system of rules. Moreover, in view of the methodological perspective that this paper adopts, the task of an inductive logician is no more than to supply these models, in order to bring out the inductive assumptions that drive analogical predictions and provide the means to manipulate these assumptions.

VII Generalizing the analogy partition

This section argues that the steps taken above can be generalized to cover analogical predictions based on explicit similarity more generally. It considers the extension of the foregoing to cases with more than two underlying predicate families. It further suggests how a problem for the model of Hesse, as noted in Maher (2001), can be solved. The solution opens up a number of interesting modelling possibilities.

Until now we have been concerned with explicit similarity between predicate families M and G, but nothing precludes the use of more than two underlying predicates, or of more cells within each partition. With the same construction we can model predictions based on explicit similarity relations that are much more complex than the above. As an example, let us say that before recording gender and marital status, we observe the type of car B in which the individual arrives, and that we distinguish between family cars, $b = 0$, vans, $b = 1$, and sports cars, $b = 2$. We may then keep track of a dependency between marital status and gender, which can on itself be made dependent on the car type. As an example, the parameter space for that partition may be

$$A = C_B \times \prod_b C_{bG} \times \left(\prod_g C_{bgM} \right), \tag{49}$$

All simplexes can again be associated with separate prediction rules, leading to an extended system of prediction rules. Note that the simplex C_B is an equilateral triangle, and that the corresponding $\lambda\gamma$ rule has three possible observation results. It will be clear that in adding further underlying predicates there are no restrictions.

As already discussed in section IV, the system of rules is suitable for expressing analogical predictions based on explicit similarity only, which means that it cannot express all possible configurations of symmetric relevance relations between the Q-predicates. To characterize the restriction on expressible relevance relations for the general case, recall first that the analogy partition always determines a certain order to the underlying predicates, such as first B, then G, and finally M. If we associate these relations with an increasing ranking number, the restriction to expressible relevance configurations may be characterised as follows: the system of prediction rules can only distinguish between the relevances of a predicate Q^q to the predicates Q^v and Q^w if the ranking numbers of the first predicate in the ranking that Q^v and Q^w do not have in common with Q^q are not the same. In other words, the system is not able to model a difference between relevances of Q^q to predicates Q^v and Q^w if the first predicate in the ranking in which the latter two differ from Q^q is the same. We may for example consider husbands more relevant to wives than to maidens. However, the system of rules cannot model these relations between the Q-predicates, because both maidens and wives differ from husbands in the first underlying predicate family in that example, namely in G.

The system of rules thus accommodates explicit similarity specifically of gender, or of marital status, but never of both. One exception to this may now be presented by slightly adapting the above example with three underlying predicates. For simplicity, the family B only concerns family cars, $b = 0$, and vans, $b = 1$. Imagine that we think that driving a van is somehow indicative of the gender, suggesting male drivers irrespective of their marital status, and further that we think family cars are indicative of the marital status, suggesting a married driver irrespective of their gender. In that case it is natural to employ the following analogy partition:

$$A = C_B \times C_{0M} \times C_{1G} \times C_{00G} \times C_{01G} \times C_{10M} \times C_{11M}. \tag{50}$$

Conditional on the individual driving a family car, we make the marital status indicative of the marital status of further family car drivers. But conditional on the individual driving a van, we make the gender indicative of the gender of further van drivers. In other words, the direction of dependence relations may vary within one analogy partition, as long as these directions are themselves conditioned on different predicates from a third family.

In the remainder of this section I illustrate one further generalisation, which uses the statistical model to capture configurations of relevance relations ρ that are not covered by the system of $\lambda\gamma$ rules. To this aim I discuss an example from Maher (2001), which reveals a shortcoming in the model of analogical predictions proposed by Hesse. Contrary to that model, the generalised statistical model can deal with the example case. It is notable that the model of Carnap and Kemeny also overcomes the difficulties of Hesse, and that their model is still more general than the model sketched here when it comes to expressible relevance relations. However, the present solution offers a number of unexplored modelling possibilities, which may eventually solve the problems with the model of Carnap and Kemeny as well.

The example of Maher concerns the predicates of being a swan X, being Australian Y, and being white Z. The indices x, y and z are 1 or 0 for the predicate being satisfied or not. Imagine that until now we have recorded, of all animals in the world, whether they are a swan and whether they are Australian. Specifically, we have observed

$$S_{\infty}^{XY} = X_1^1 \cap Y_1^0 \cap X_2^1 \cap Y_2^1 \cap \left(\bigcap_{i=3}^{\infty} X_i^{x_i} \cap Y_i^{y_i} \right), \tag{51}$$

the sequence of observations of all animals with respect to being a swan and being Australian, the first animal in the sequence being a non-Australian swan, the second an Australian swan. The challenge is to define a probability over the algebra $\mathcal{Q}$, or equivalently $\mathcal{R}$, for which

$$p(Z_2^1 | Z_1^1 \cap S_{\infty}^{XY}) > p(Z_2^1 | Z_1^0 \cap S_{\infty}^{XY}). \tag{52}$$

That is, we want the whiteness of a non-Australian swan to be relevant to the probability of the whiteness of an Australian swan. The fact that we already know the proportions of Australian and non-Australian swans and non-swans should have nothing to do with this. But unfortunately the model of Hesse cannot accommodate such a relevance.

It turns out that the above inequality can be derived by employing a restricted prior over an analogy partition. For simplicity, use the parameter space $A = C_{XY} \times C_{00Z} \times C_{01Z} \times C_{10Z} \times C_{11Z}$, which has a single, three-dimensional simplex C_{XY} for the combined predicates X and Y on being a swan and being Australian. In this tetrahedron space, conditioning on S_{∞}^{XY} forces all probability to be concentrated on one point hypothesis within the simplex C_{XY}, associated with the actual relative frequencies α_{xy} of the observations in S_{∞}^{XY}. Subsequent observations Z_i^z on being white therefore only influence the probability over the remaining spaces C_{xyZ} for each of the values xy. Now the challenge is to establish the relevance of observations of being white within the category $xy = 10$, concerning non-Australian swans, to the probability of animals being white in the category $xy = 11$, concerning Australian swans.

In other words, we must somehow couple the probability assignment over the simplex C_{11Z} to the assignment over C_{10Z}.

One way of doing this is by restricting the probability assignment to a specific subspace of the hypotheses space A, defined by $\alpha_{10z} = \alpha_{11z}$. Effectively, the marginal probability assignments over the two simplexes C_{10Z} and C_{11Z} are then identified, so that adapting the probability over C_{10Z} for the observation Z_i^1, given that $X_i^1 \cap Y_i^0$, implicitly changes the probability over C_{11Z} as well. In other words, finding a non-Australian swan to be white is immediately relevant for the probability of Australian swans being white. The probability over the remaining spaces on non-swans, $C_{01Z} \times C_{00Z}$, can be chosen freely.

Within the hypotheses space on swans, in which all probability is restricted to $\alpha_{10z} = \alpha_{11z}$, we may again choose a Dirichlet distribution. For the resulting predictions, this means that swans, both Australian and non-Australian, are collected in the same $\lambda\gamma$ prediction rule on observations of being white, Z:

$$p(Z_{n+1}^z | S_n^Z \cap S_\infty^{XY}) = \frac{n_{Zz}^{11} + n_{Zz}^{10} + \lambda_Z^1 \gamma_{Zz}^1}{n_Z^{11} + n_Z^{10} + \lambda_Z^1}. \tag{53}$$

For this prediction to be applicable, we must have that $S_\infty^{XY} \subset X_{n+1}^1$, meaning that animal $n+1$ is indeed a swan. Predictions for the case in which $S_\infty^{XY} \subset X_{n+1}^0$ are determined by the probability over the space on non-swans, $C_{01Z} \times C_{00Z}$. Note that S_n^Z denotes the sequence of observations of the first n animals with respect to Z. As in the above, n_Z^{11} and n_Z^{10} are defined as the numbers of Australian and non-Australian swans in the sequence S_n^Z, and n_{Z1}^{11} and n_{Z1}^{10} as the numbers of Australian and non-Australian white swans in that sequence. The above rule further uses the abbreviations $\lambda_Z^1 = \lambda_Z^{11} = \lambda_Z^{10}$ and $\gamma_{Zz}^1 = \gamma_{Zz}^{11} = \gamma_{Zz}^{10}$. These parameters are the same for the simplexes C_{11Z} and C_{10Z}, because they are determined by one and the same probability assignment.

It can be checked that the inequality (52) is indeed satisfied in this model. But it must also be conceded that the above model has its shortcomings when it comes to the expressibility of relevance relations. Moreover, there are many more possibilities with non-factorisable priors over the partition $\mathcal{A}$ that have not been investigated in the above. A much more detailed study of analogical predictions based on such priors is necessary in order to make any more general claims on its relative merits and defects. Accordingly, the main aim of the above example is to suggest that the use of statistical models deserves further attention.

VIII Conclusion

This paper presents a system of $\lambda\gamma$ rules that models analogical predictions based on analogy by explicit similarity of gender. After presenting an example of such similarity, the paper shows how it translates to a specific subset of relevance relations between predicates in the aggregated family Q: the relevance of the predicate Q^{2g+m} for the predicate $Q^{2g+m'}$, which has predicate G^g in common with Q^{2g+m}, differs from its relevance for the two predicates $Q^{2g'+m}$ and $Q^{2g'+m'}$, that do not have G^g in common with Q^{2g+m}.

After presenting a system of rules that indeed models these relevance relations, I provide the Bayesian model that underlies the system. It is shown that analogy hypotheses treat observations M^m separately for the earlier observation G^g with $g = 0, 1$, by defining separate relative frequencies for them, and associating these frequencies with separate dimensions in the parameter space A. By assuming the prior over this space to be a product of Dirichlet marginals, the system of $\lambda\gamma$ rules can be derived. The paper ends with some generalizations on the proposed system of rules.

Future research will explore the possibilities of the Bayesian model that underlies the system of rules. It will turn out that this model provides the setting for a completely general model of analogical predictions based on symmetric inductive relevance relations, by employing non-factorisable probability functions over A. More generally, on the level of research programmes, I take it to be an important advantage of the present model that it seeks to integrate the rather isolated discussion on analogical predictions within Carnapian inductive logic in a wider framework of Bayesian statistical inference.

Acknowledgements

I want to thank Roberto Festa, Theo Kuipers, and Patrick Maher for comments and discussions on earlier versions of this article. The inaccuracies, and possibly more serious defects of the position defended in this article are of course my responsibility.

References

[1] Bovens, L. and Hartmann, S. 2004, *Bayesian Epistemology*, Oxford University Press, Oxford.

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

[3] Carnap, R.: 1980, 'A Basic System of Inductive Logic, Part II', in Jeffrey, R.C. (ed.), *Studies in Inductive Logic and Probability*, Vol.2, University of California Press, Berkeley, pp. 7-150.

[4] Carnap, R. and Stegmüller, W.: 1959, *Induktive Logik und Warscheinlichkeit*, Springer, Wien, pp. 242-52.

[5] De Finetti, B.: 1964, 'Foresight: its Logical Laws, its Subjective Sources' in Kyburg, H. E. and Smokler, H. E. (eds.) *Studies in Subjective Probability*, John Wiley and Sons, New York.

[6] Finetti, B. de: 1974, *Theory of Probability* Vol. 1, John Wiley, New York.

[7] Di Maio, M. C.: 1995, 'Predictive Probability and Analogy by Similarity in Inductive Logic', *Erkenntnis* **43**, pp. 369 - 394.

[8] Festa, R.: 1993, *Optimum Inductive Methods*, Kluwer Academic Publishers, Dordrecht.

[9] Festa, R.: 1997 "Analogy and Exchangeability in Predictive Inferences", *Erkenntnis* **45**, pp. 89-112.

[10] Howson, C. and Urbach, P.: 1996, *Scientific Reasoning: the Bayesian Approach*, Open Court Publishing Company, Chicago.

[11] Jeffrey, R.: 1984, *Probability and the Art of Judgement*, Cambridge University Press, Cambridge.

[12] Kuipers, T.A.F.: 1984, 'Two Types of Inductive Analogy by Similarity', *Erkenntnis* **21**, pp. 63-87.

[13] Kuipers, T.A.F.: 1988, 'Inductive Analogy by Similarity and Proximity', in Helman, D. (ed.) *Analogical Reasoning*, Kluwer Academic Publishers, Dordrecht, pp. 271-99.

[14] Maher, P.: 2000, 'Probabilities for Two Properties', *Erkenntnis* **52**, pp. 63-91.

[15] Maher, P.: 2001, 'Probabilities for Multiple Properties: the models of Hesse and Carnap and Kemeny', *Erkenntnis* **55**, pp. 183-216.

[16] Niiniluoto, I.: 1981, 'Analogy and Inductive Logic', *Erkenntnis* **21**, pp. 1-34.

[17] Niiniluoto, I.: 1988, 'Analogy by Similarity in Scientific Reasoning', in Helman, D. (ed.) *Analogical Reasoning*, Kluwer Academic Publishers, Dordrecht, pp. 271-99.

[18] Norton, J. D.: 2003, 'A Material Theory of Induction', Philosophy of Science **70**, 647-670.

[19] Skyrms, B.: 1993, 'Analogy by Similarity in Hyper-Carnapian Inductive Logic', in J. Earman, A.I. Janis, G. Massey, and N. Rescher (eds.), *Philosophical Problems of the Internal and External Worlds*, University of Pittsburgh Press, Pittsburgh, pp. 273-282.

[20] Spohn, W.: 1981, 'Analogy and Inductive Logic: a Note on Niiniluoto', *Erkenntnis* **21**, pp. 35-52.