A General Model for Analogical Predictions

This chapter presents a general model for exchangeable analogical predictions, based on inductive relevance between four predicates. It extends the model for analogy by explicit similarity, which was dealt with in the preceding chapter. It is first shown that the predictions of this model may be captured in a more general statistical framework for exchangeable predictions. This is then used to define a model that is able to incorporate all possible symmetric inductive relevance relations. However, not all aspects of analogical predictions find a natural formalisation in the model. The chapter therefore only offers a partial explication of the model alongside some numerical approximations. At the end the model of this chapter is related to some other models in the literature.

It is not necessary to read chapter 4, or any of the other chapters, before starting on this one. But it may be noted that chapter 4 provides a different perspective on very similar inductive schemes. There is considerable overlap in the technical parts of the chapters. Reading the preceding chapter may therefore be helpful in coming to a more complete understanding of the schemes discussed in this chapter.

5.1 Introduction

Analogical predictions deviate from standard Carnapian predictions, as for example in Carnap (1952), because analogical predictions incorporate considerations of inductive relevance between predicates. This section discusses a general scheme for capturing such relevance relations. After that I make explicit the kind of analogical predictions that this chapter deals with, and contrast them with the analogical predictions of the preceding chapter.

5.1.1 Inductive relevance

Party centre example. Consider a marketing manager of a party centre who makes predictions on visitors concerning their gender and marital status. There are four aggregated predicates: bachelors, husbands, maidens, and wives. Let us say that the manager does not know the party centre yet, but that she has
some general knowledge of party centres. Specifically, she knows that men like 
to hang out in a party centre together, so that recording a bachelor is positively 
relevant to recording a husband and vice versa. She further knows that married 
couples often go out together, and also that, while husbands tend to bring their 
bachelor friends if they go out with their wives, the wives tend not to invite 
their maiden friends. Therefore, while husbands and wives are strongly relevant 
to each other, as husbands and bachelors are, maidens and wives bear a much 
weaker relevance relation. The challenge of this chapter is to find a model for 
inductive predictions that incorporates complex configurations of such relevance 
relations.

A vector of relevance relations. Consider the notion of inductive relevance itself. 
We can formalise the above predicates as $Q^q_i$, with the numbers $q = 0, 1, 2, 3$ 
associated with the predicates bachelor, husband, maiden and wife respectively. 
In terms of these predicates, the example has it that $Q^1$ is more relevant to $Q^0$ 
than to $Q^2$. This inductive relevance means that the observation that individual 
i has predicate $Q^1$, or $Q^1_i$ for short, is more favourable to the probability of the 
observation $Q^0_{i+1}$ than to that of $Q^2_{i+1}$. Note that this may obtain quite inde-
pendently of the purely inductive effect that observation $Q^1_i$ makes observation 
$Q^1_{i+1}$ more probable.

The relevance of $Q^q$ to $Q^w$ may be expressed in a function $\rho(q, w)$, where $q$ 
and $w$ denote predicate numbers. With this relevance function, we may specify 
in general terms what it means for $Q^q$ to be more relevant to $Q^w$ than to $Q^v$. 
Denoting the probability of observation $Q^q_i$ with the function $p$, we can write 

$$\rho(q, w) > \rho(q, v) \Rightarrow \frac{p(Q^w_{i+1}|E_{i-1} \cap Q^q_i)}{p(Q^w_{i+1}|E_{i-1})} > \frac{p(Q^v_{i+1}|E_{i-1} \cap Q^q_i)}{p(Q^v_{i+1}|E_{i-1})}. \quad (5.1)$$

It must be noted that the foregoing is not the only possible expression of inductive 
relevance. For a review of possible relations between inductive methods and 
the relevance function, see Festa (1997). The above is qualitatively equivalent 
to $K_{>G}$ inductive methods in his terminology.

This chapter deals with complex configurations of relevance relations, like 
those exemplified in the party centre example. By means of the above relevance 
function $\rho(v, w)$, its aim can be made a bit more specific. First, as in the 
preceding chapter, I restrict attention to symmetric relations:

$$\rho(v, w) = \rho(w, v). \quad (5.2)$$
Because the relevance relations are symmetric, the set of possible configurations of relevance between predicates may be captured in the following space:

\[ \rho = (\rho(0, 1), \rho(2, 3), \rho(0, 2), \rho(1, 3), \rho(1, 2), \rho(0, 3)). \]  

(5.3)

The more specific aim of this chapter is to provide a model of inductive predictions for all the relevance configurations that are represented in this space.

Some disclaimers. It may be remarked immediately that some important aspects of inductive relevance are not expressed in the representation \( \rho \). First of all, expression (5.1) does not yet provide a meaning for the sizes of \( \rho \). Until now the characterisation is entirely qualitative, only providing an interpretation for the ordering of the sizes. To a certain extend the numerical values of \( \rho \) are given a further interpretation below. Second, because I am considering exchangeable predictions, these predictions will converge onto the actual relative frequencies of the predicates \( Q^q \). The analogy effects will therefore diminish with the accumulation of observations. But the foregoing does not specify exactly how the effects will diminish, or even the overall rate at which this happens. This aspect is simply not captured in the representation \( \rho \). It may be that the rates are different for the different relevance relations, so that the ordering in these relations varies with the number and the nature of the observations. However, as will be argued below, in the present model the analogy effects diminish at the same rate. The relevance ordering therefore remains intact.

5.1.2 Aim of this chapter

The model for explicit similarity. It is instructive to relate it to the model of the preceding chapter, which concerned analogical predictions based on explicit similarity relations. The aggregate predicate family \( Q \) on bachelors, husbands, maidens and wives is built up from separate predicate families on gender, \( G \), and marital status, \( M \). We can write

\[ G^g = Q^{2g} \cup Q^{2g+1}, \]  

(5.4)

\[ M^m = Q^m \cup Q^{2+m}, \]  

(5.5)

where \( G^g \) means male for \( g = 0 \) and female for \( g = 1 \), while \( M^m \) means not married for \( m = 0 \) and married for \( m = 1 \). The idea of the model for explicit similarity is that predictions on predicates \( Q^q \) may be written down in terms of predictions for the separate families \( G \) and \( M \):

\[ p(Q_{n+1}^q | E_n) = \frac{n_G g + \lambda_G g g}{n_G + \lambda_G} \times \frac{n_M m + \lambda_M g m}{n_M + \lambda_M}. \]  

(5.6)
Here $n_G = n$ is the total number of individuals in the preceding observations $E_n$, $n_{Gg} = n_{Gg}^2$, the number of individuals with gender $g$, and $n_{Gm}^2$, the number of individuals with gender $g$ and marital status $m$.

Recall that every point in the vector space $\rho$ represents a different configuration of symmetric inductive relevance relations. The models based on explicit similarity can now be captured by a specific subset of configurations in this vector space, namely:

$$\rho = \langle \rho_{G0}, \rho_{G1}, \rho_{G}, \rho_{G}, \rho_{G}, \rho_{G} \rangle.$$ (5.7)

The relevances $\rho_{Gg} = \rho(2g, 2g + 1)$, the so-called intra-gender relevances, may be chosen independently, and the inter-gender relevances $\rho_{G}$ must all be chosen equal. They are connected to the parameters in the above prediction rule according to

$$\lambda_M^g = \rho_{Gg} \gamma_{Gg} N,$$ (5.8)
$$\lambda_G = \rho_G N.$$ (5.9)

The model of this chapter is a generalisation of the model for explicit similarity: every component of $\rho$ can in this chapter be determined independently.

**Position of the present chapter.** The ideas in this chapter are strongly connected to earlier chapters. As indicated, the present model relies heavily on the model for explicit similarity, as it is discussed in the preceding chapter. The results of that chapter are here used uncritically. Furthermore, just as the preceding one, this chapter employs statistical hypotheses for generating inductive predictions. That is, the model first incorporates observations on predicates into a probability assignment over statistical hypotheses, and then employs the assignment over these hypotheses to derive predictions for new observations. As discussed in chapter 3, the use of hypotheses offers control over the assumptions underlying inductive predictions. The present chapter illustrates this. It shows how transformations between hypotheses spaces enable us to define priors that are otherwise hard to find.

The plan of this chapter is as follows. Sections 5.2 and 5.3 introduce observational algebras for the above $Q$-predicates and for the predicates such as $G$ and $M$ that underlie them, and defines the predictions based on hypotheses concerning these respective predicates. Section 5.4 elaborates on the relation between $Q$-predicates and underlying predicates. Specifically, it shows how the prior probability over hypotheses concerning underlying predicates that encodes
5.2. HYPOTHESES SCHEMES FOR $Q$-PREDICATES

Observational algebra. The expression $Q^q_i$ refers to the observation that individual $i$ has predicate $Q^q$. To characterise inductive predictions, let me represent these observations in terms of a so-called observational algebra. Let $K$ be the set of possible values for $q$, so that in the case of the party centre $K = \{0, 1, 2, 3\}$. The infinite product $K^\omega$ is the space of all infinite sequences $e$ of such values:

$$e = q_1 q_2 q_3 \ldots$$

(5.10)

The observational algebra, denoted $Q$, consists of all possible subsets of the space $K^\omega$. If we denote the $i$-th element in a series $e \in K^\omega$ with $e(i)$, we can define an observation $Q^q_i$ as an element of the algebra $Q$ as follows:

$$Q^q_i = \{ e : e(i) = q \}.$$  

(5.11)

Note that there is a distinction between the observation $Q^q_i$ and the result of an observation $q$. The values, represented with small letters, are natural numbers. The observations, denoted with large letters, are elements of the algebra $Q$.

In the same way we can define an element in the algebra that represents a finite sequence of observations. If we define the ordered $n$-tuple $e_n = (q_1, q_2, \ldots, q_n)$ and $q_i$ as the $i$-th element therein, we can write

$$E^e_n = \{ e : \forall i \leq n \ (e(i) = q_i) \}.$$  

(5.12)
I normally suppress reference to the \( n \)-tuple \( e_n \). The observations and sequences of observations are related to each other according to

\[
E_n \cap Q^q_{n+1} = E_{n+1}
\]  

(5.13)

where \( e_{n+1}(n + 1) = q \). Finally, for any sequence \( e_n \) we can write down, for all numbers \( q < 4 \), the number of times it occurs within the sequence. These numbers are in the following denoted with \( n_{Qq} \). Since the total number of observations \( n_Q = n = \sum_q n_{Qq} \), the numbers \( n_{Qq} \) together define the observed relative frequencies \( \frac{n_{Qq}}{n_Q} \) of the results \( q \).

We can now define a probability function \( p \) over the algebra \( Q \). The probabilities of observations \( Q^q_{n+1} \) and \( E_n \) can then be interpreted as predictions. An important matter is how these predictions depend on observations: if the series \( e_n \) is observed, this must somehow change the predictions over the observations. I here assume that the predictions upon observing \( e_n \) are expressed by the original probability function \( p \) conditional on the observations \( E_n \), denoted \( p(\cdot|E_n) \).

This dependence of predictions on observations is known as Bayesian conditioning. In the following, the initial probability is called the prior probability, and the conditional one the posterior.

**Bernoulli hypotheses and exchangeable predictions.** The schemes of this chapter employ partitions of statistical hypotheses to define the probability function \( p \). A partition is a collection \( \mathcal{B} = \{H_\theta\}_{\theta \in \mathcal{B}} \) in which the hypotheses \( H_\theta \) are mutually exclusive and jointly exhaustive possibilities. We can associate these hypotheses with elements of the algebra \( Q \), but for present purposes the less strict characterisation suffices. We can define predictions \( p(Q^q_{n+1}|E_n) \) with the law of total probability over the partition:

\[
p(Q^q_{n+1}|E_n) = \int_\mathcal{B} p(H_\theta|E_n)p(Q^q_{n+1}|H_\theta \cap E_n) \, d\theta.
\]  

(5.14)

The probability over the hypotheses is determined by the probability density \( p(H_\theta|E_n) \). The terms \( p(Q^q_{n+1}|H_\theta \cap E_n) \) are called the likelihoods on the hypotheses \( H_\theta \), which are defined for observations \( Q^q_{n+1} \). The prediction is obtained by weighing these likelihoods with the posterior probability over the hypotheses.

Apart from the likelihoods, the dependence of the predictions on observations are reflected in the probability assignment over the hypotheses. This probability may be determined by means of Bayesian conditioning,

\[
p(H_\theta|E_{i+1}) \, d\theta = \frac{p(Q^q_{i+1}|H_\theta \cap E_i)}{p(Q^q_{i+1}|E_i)} p(H_\theta|E_i) \, d\theta,
\]  

(5.15)
where \( e_{i+1}(i+1) = q \). Note that the denominator \( p(Q_{i+1}^n|E_i) \) can be rewritten with equation (5.14). The posterior probability over the hypotheses \( p(H_\theta|E_n)d\theta \) can thus be determined recursively by the prior probability assignment \( p(H_\theta)d\theta \), and the likelihoods \( p(Q_{i+1}^n|H_\theta \cap E_i) \) for all times \( 0 \leq i < n \). With the likelihoods \( p(Q_{n+1}^n|H_\theta \cap E_n) \) we can then determine the predictions.

This chapter employs specific statistical hypotheses \( H_\theta \), which have the following likelihoods for the observations \( Q_{i+1}^n \):

\[
p(Q_{i+1}^n|H_\theta \cap E_i) = \theta_q. \tag{5.16}
\]

The domain of the 4-tuple \( \theta \) is a simplex, \( \sum \theta_q = 1 \). Note also that the likelihoods are independent of the earlier observations \( E_i \). The posterior likelihoods are thus identical to the prior likelihoods. Finally, the predictions resulting from the partition \( B \) are exchangeable, and by De Finetti’s representation theorem every exchangeable prediction rule can be captured by a prior probability over the hypotheses, \( p(H_\theta)d\theta \).

**Carnapian predictions.** One specific prior probability assignment must be given separate attention. If we assume the prior density function over the simplex to have a Dirichlet form,

\[
p(H_\theta) \sim \prod_q \theta_q^{(c_q-1)}, \tag{5.17}
\]

then the resulting predictions are of the form of Carnapian rules \( p_{\lambda\gamma} \):

\[
p(Q_{n+1}^n|E_n) = \frac{n_{Qq} + \lambda Q\gamma Qq}{n_Q + \lambda Q} = pr_{\lambda\gamma}(n_{Qq}, n_Q). \tag{5.18}
\]

The values of the parameters \( \lambda_Q \) and \( \gamma_Qq \) are determined by the exponents \( c_q \) in the Dirichlet density according to \( \lambda = \sum c_q \) and \( \gamma_q = c_q/\lambda \). In this chapter I restrict attention to natural numbers \( c_q \). Finally, the numbers \( n_{Qq} \) and \( n_Q \) are as defined in the preceding section.

### 5.3 Schemes using underlying predicates

This section presents the algebra for the underlying predicates \( G \) and \( M \). It further introduces the hypotheses partition \( A \) associated with this algebra, and argues that this partition leads to the system of \( \lambda\gamma \) rules of section 5.1. This section shows considerable overlap with section 4.6, but it is somewhat more general.

**Algebra for underlying predicates.** Let me define an observation algebra for the underlying predicates \( G \) and \( M \), as introduced in section 5.1.1. Recall the
indices of these predicates, \( g, m \in \{0, 1\} \). With \( L_{GM} \) as the set of ordered pairs \( \langle g, m \rangle \) we can define the space \((L_{GM})^\omega \) of all infinitely long ordered sequences \( u \) of such index pairs:

\[
u = g_1 m_1 g_2 m_2 g_3 m_3 \ldots
\]

(5.19)

We can then identify all infinite strings of observations \( e \in K^\omega \) with a unique infinite string \( u \in (L_{GM})^\omega \):

\[
e(i) = 2g_i + m_i,
\]

(5.20)

\[
u(t) = \begin{cases} g_i & \text{if } t = 2i - 1, \\ m_i & \text{if } t = 2i. \end{cases}
\]

(5.21)

So for every odd index \( t \), the number \( u(t) \) concerns an observation of predicate \( G \), and for every even index \( t \) the number \( u(t) \) concerns \( M \). Thus every sequence \( e \) is mapped onto a unique sequence \( u \), and every such \( u \) can be traced back to a corresponding sequence \( e \).

Two things must be remarked on this translation of \( Q \)-predicates to underlying predicates. First, note that the order of the underlying predicates \( G \) and \( M \) is fixed in the definition of the ordered pairs \( L_{GM} \). However, we may just as well consider the set \( L_{MG} \), and define a space of infinite sequences \( u' \) on the basis of that. Moreover, the predicates \( G \) and \( M \) do not exhaust the possible pairwise combinations of \( Q \)-predicates. We can also employ a third partitioning of the \( Q \)-predicates:

\[
W^w = Q^{1-w} \cup Q^{2+w}.
\]

(5.22)

As a slightly contrived interpretation, imagine that it is a custom of the people featuring in the example that bachelors are given a traditional wedding ring at their 18th birthday. This ring is a sign that the bachelor has reached the age at which he is allowed to propose to a maiden, and it serves as a present to the bride at the wedding ceremony. Therefore, people who are in possession of this traditional ring are either bachelors or wives, and people who are not are either maidens or husbands. The important thing here is that translations of sequences \( e \) into sequences in a space based on \( L_{WG} \) or on some other combination with the family \( W \), may be considered just as well.

We can now define the algebra \( R_{GM} \) for observations of predicate families \( G \) and \( M \) in the space \((L_{GM})^\omega \), in the same way as we defined the algebra \( Q \):

\[
G^g_i = \{ u \in (L_{GM})^\omega : u(2i - 1) = g \},
\]

(5.23)

\[
M^m_i = \{ u \in (L_{GM})^\omega : u(2i) = m \}.
\]

(5.24)
5.3. SCHEMES USING UNDERLYING PREDICATES

The sets $G_i^g \cap M_i^m$ contain all those infinitely long sequences $u \in (L_{GM})^\omega$ that have the number $g$ and $m$ in the positions $2i-1$ and $2i$ respectively. We can therefore also translate

$$Q_i^{(2g+m)} = G_i^g \cap M_i^m, \quad (5.25)$$
$$E_n^m = S_n^m = \bigcap_{i=1}^n G_i^g \cap M_i^m. \quad (5.26)$$

In this way there is a complete mapping of the elements $Q_i^g$ and $E_n$ in $Q$ onto elements of the algebra $R_{GM}$.

**Hypotheses for underlying predicates.** We may define inductive predictions concerning $Q$-predicates by providing a probability function $p$ over the algebra of underlying observations, $R_{GM}$. Again we can employ a partition of hypotheses $H_\alpha$ with the parameter space $\alpha \in A_G$, so that $A_G = \{H_\alpha\}_{\alpha \in A_G}$. The hypotheses $H_\alpha$ concern observations $G_i^g$ and $M_i^m$, that is, they provide likelihoods for these observations. We may choose

$$p(G_{i+1}^g | H_\alpha \cap S_i) = \alpha_{G^g}, \quad (5.27)$$
$$p(M_{i+1}^m | H_\alpha \cap G_{i+1}^g \cap S_i) = \alpha_{G^m}. \quad (5.28)$$

These likelihoods do not depend on observations concerning other individuals: the parameters $\alpha$ do not depend on $S_n$. However, every hypothesis does have separate likelihoods for observations $M_i^m$ conditional on either $G_i^0$ or $G_i^1$.

We can use the hypotheses $H_\alpha$ to generate predictions over the underlying predicate families $G$ and $M$, just as we used hypotheses $H_\theta$ for direct predictions of the family $Q$. Since the algebra $R_{GM}$ determines that we observe $G_{i+1}^g$ before observing $M_{i+1}^m$, all relevant likelihoods are in this way defined. The main difference with the above discussion is in the parameter space $A_G$. Instead of a single simplex $B$ with $\sum \theta_q = 1$, we now have a Cartesian product of three simplexes $A_G = B_G \times B_0M \times B_1M$, with components:

$$\sum_g \alpha_{G^g} = 1, \quad \sum_m \alpha_{G^0m} = 1, \quad \sum_m \alpha_{G^1m} = 1. \quad (5.29)$$

Since in the example we have $g, m = 0, 1$, we simply have $\alpha_{G^1} = 1 - \alpha_{G^0}$ and $\alpha_{G^01} = 1 - \alpha_{G^0}$.

**Carnapian rules for underlying predicates.** As in the foregoing, the probability density over the hypotheses space determines the eventual predictions that derive from the hypotheses scheme. And indeed, if we assume a Dirichlet density
over each separate simplex component of the parameter space $A_G$,

$$p(H_n) \sim \prod_g \left( \alpha_{Gg}^{(a_{Gg} - 1)} \times \prod_m \alpha_{Ggm}^{(a_{Ggm} - 1)} \right), \quad (5.30)$$

we can derive Carnapian $\lambda\gamma$ rules for the predicate families $G$ and $M$ separately:

$$p(G^g_{n+1}|S_n) = \frac{n_{Gg} + \lambda_{Gg} \gamma_{Gg}}{n_G + \lambda_G}, \quad (5.31)$$

$$p(M^m_{n+1}|S_n \cap G^g_{n+1}) = \frac{n^g_{Mm} + \lambda^g_{Mm} \gamma^g_{Mm}}{n^g_M + \lambda^g_M}. \quad (5.32)$$

The different dimensions in the parameter space are responsible for the independent prediction rules over $G^g_{n+1}$, and over $M^m_{n+1}$ conditional on $G^g_{n+1}$ and $G^g_{n+1}$ respectively. Further, the numbers $n_G$, $n_{Gg}$, $n^g_M$ and $n^g_{Mm}$ are as indicated in section 5.1.

To complete the statistical underpinning of the system of $\lambda\gamma$ rules, recall the relation between the observations of families $G$ and $M$ on the one hand, and the observations of family $Q$ on the other. We can write

$$p(Q^q_{n+1}|E_n) = p(G^g_{n+1}|S_n) \times p(M^m_{n+1}|S_n \cap G^g_{n+1}), \quad (5.33)$$

and thus arrive at the model for analogical predictions presented in equation (5.6). As indicated in section 5.1 and in the preceding chapter, this model allows us to express a specific subset of inductive relevance configurations.

Finally, let me provide the relations between the prior probability assignment over $A_G$ and the parameters in the above prediction rules. As in the case of the prediction rule over $Q$-predicates, the exponents in the Dirichlet prior are directly related to these parameters:

$$\lambda_G = \sum_g a_{Gg} \quad \gamma_{Gg} = \frac{a_{Gg}}{\lambda_G}, \quad (5.34)$$

$$\lambda^g_M = \sum_m a_{Ggm} \quad \gamma^g_{Mm} = \frac{a_{Ggm}}{\lambda^g_M}. \quad (5.35)$$

In choosing the Dirichlet priors over the simplex components of $A_G$ we thus have separate command over the three Carnapian $\lambda\gamma$ rules in the system.

### 5.4 Transformations between partitions

The above presents a partition of statistical hypotheses for exchangeable predictions on $Q$-predicates. It also provides a specific partition for statistical
5.4. TRANSFORMATIONS BETWEEN PARTITIONS

hypotheses that, with a similar prior, results in exchangeable analogical predictions for explicit similarity. In this section we translate the prior over this latter partition into a prior over the general partition. This prepares for the general model of the next section.

5.4.1 COORDINATE TRANSFORMATIONS

Equivalence of B and A. Consider the two partitions B and A_G. Recall that the parameter components of the partition B, denoted \( \theta_q \), are the likelihoods for the observations \( Q_q \), as expressed in (5.16). These likelihoods determine the nature of the partition: if we provide a prior probability over it, the predictions are determined. But it can further be noted that the partition A_G indirectly determines likelihoods for the \( Q \)-predicates as well:

\[
p(Q_{i+1}^{2g+m} | H_\alpha \cap S_i) = p(G_{i+1}^g \cap M_{i+1}^m | H_\alpha \cap S_i) = p(G_{i+1}^g | H_\theta \cap S_i) p(M_{i+1}^m | H_\theta \cap S_i \cap G_{i+1}^g) = \alpha_{Gg} \alpha_{Ggm}. \tag{5.36}
\]

This determines the update operation over A_G that corresponds to an update operation with \( Q_{i+1}^{2g+m} \) over B.

Every hypotheses \( H_\alpha \) may now be identified with a hypothesis \( H_\theta \) according to the set of transformation rules determined by the above equivalence:

\[
\theta_{2g+m} = \alpha_{Gg} \alpha_{Ggm}. \tag{5.37}
\]

Note that there are 4 components of \( \theta \) that have to comply to 1 normalisation condition, so that B has 3 degrees of freedom. Since there are 6 components of \( \alpha \) that have to comply to 3 normalisation conditions, the number of degrees of freedom in A_G is also 3. The mapping of equation (5.37) is in fact a bijection: for any hypothesis \( H_\theta \) there is a unique \( H_\alpha \) that has the same likelihoods for the observations \( Q_{i+1}^q \). The partitions B and A_G are therefore essentially the same. This also means that the results on exchangeability and convergence, which may be proved for partition B, hold for the partition A_G as well.

On the other hand, the structures of the parameter spaces B and A_G are certainly not identical. And as suggested above, this difference can be employed to access distributions over the hypotheses in B that are very difficult to come up with, or to investigate properties of, using the parameter space B itself. The access is provided by first defining the prior probability over the equivalent partition A_G, employing the characteristics of the prediction rules defined for that
partition, and by subsequently transforming this prior into one over the partition \( B \). The prior over \( B \) that is obtained in this way will result in the very same predictions as those derivable from \( A_G \), which, as may be recalled, incorporate analogical effects of explicit similarity. For the purpose of this chapter it is most significant that the translation reveals the characteristics of a prior probability over \( B \) that are responsible for the kind of analogical predictions derivable from \( A_G \). Eventually this leads the way to the definition of a class of priors over \( B \) that incorporates all possible inductive relevance relations.

Transformation rules and Jacobian. Before doing that, let me describe the transformation for the case of the predicate families \( Q, G \) and \( M \). As for the probability function itself, we can employ the following transformation relations between the components of \( \alpha \) and \( \theta \), which can be derived from the transformation equation (5.37):

\[
\begin{align*}
\alpha_{G0} &= \theta_0 + \theta_1, \\
\alpha_{G00} &= \frac{\theta_0}{\theta_0 + \theta_1}, \\
\alpha_{G10} &= \frac{\theta_2}{\theta_2 + \theta_3}, \\
\alpha_{G11} &= \frac{\theta_3}{\theta_2 + \theta_3}.
\end{align*}
\]  

(5.38)

For any density function \( p(H) \) over \( A_G \), we can simply write all the components of \( \alpha \) as these fractions of components of \( \theta \). However, in order to make up for the change of the space itself, we must multiply the resulting function of components of \( \theta \) with the so-called Jacobian, the determinant of the transformation matrix. This method is described in any standard textbook on vector calculus, for instance Marsden (1988).

As it turns out, it is simpler to calculate the Jacobian \( J^{-1}(\alpha) \) for the inverse transformation of \( B \) to \( A_G \) first, and to derive the form of \( J(\theta) \) from that. It is further simpler to employ only the three free parameter components for the space \( A_G \):

\[
\begin{align*}
\alpha_{G} &= \alpha_{G1} = 1 - \alpha_{G0}, \\
\alpha_{0M} &= \alpha_{01} = 1 - \alpha_{00}, \\
\alpha_{1M} &= \alpha_{11} = 1 - \alpha_{10}.
\end{align*}
\]  

(5.39)

For the space \( B \) we can simply take the \( \theta_q \) with \( q = 1, 2, 3 \). The simpler transformation rules then are

\[
\begin{align*}
\theta_1 &= (1 - \alpha_{G})\alpha_{0M}, \\
\theta_2 &= \alpha_{G}(1 - \alpha_{1M}), \\
\theta_3 &= \alpha_{G}\alpha_{1M}.
\end{align*}
\]  

(5.40)
These transformations are again essentially the same as the transformation equation in (5.37).

The Jacobian $J^{-1}(\alpha)$ is now given by the determinant of the transformation matrix. The $q$-th row in this matrix consists of the partial derivatives of $\theta_q$ to the components of $\alpha$, in the order $\alpha_G$, $\alpha_{0M}$, and $\alpha_{1M}$. The Jacobian is thus given by

$$J^{-1}(\alpha) = \det \begin{pmatrix} -\alpha_{0M} & 1 - \alpha_G & 0 \\ 1 - \alpha_{1M} & 0 & -\alpha_G \\ \alpha_{1M} & 0 & \alpha_G \end{pmatrix}, \quad (5.41)$$

Writing out the determinant we find

$$J^{-1}(\alpha) = \alpha_G(1 - \alpha_G).$$

This factor makes up for the change of the infinitesimal volumes $d\alpha$ into $d\theta$ during the transformation.

### 5.4.2 Transformed probability models

**Explicit analogy in terms of $Q$-predicates.** Recall that for the system of $\lambda\gamma$ rules, the probability over the space $A_G$ is given by the density of equation (5.30). The density over the space $B$ is determined by the transformation rules of equation (5.38) and the Jacobian (5.42), and thus given by

$$p(H_\theta) \sim (\theta_0 + \theta_1)^{r_{01}} \times (\theta_2 + \theta_3)^{r_{23}} \times \prod_{g,m} \theta^{a_G_{g,m} - 1}. \quad (5.43)$$

For the exponents of the cross-terms, $r_{2g,2g+1}$, we have

$$r_{2g,2g+1} = a_{Gg} - a_{Gg0} - a_{Gg1}. \quad (5.44)$$

This prior probability over the space $B$ results in exactly the same predictions over the $Q$-predicates as can be derived from the corresponding prior over the space $A_G$, which are expressed in the predictions (5.6).

It can be noted immediately that the prior over $B$ deviates from a Dirichlet prior because of the cross-terms $(\theta_{2g} + \theta_{2g+1})^{r_{2g,2g+1}}$. These terms are responsible for the analogical effects in the predictions. Recall that the exponents in the above density are related to the system of prediction rules according to $\lambda_G \gamma G_g = a_{Gg}$ and $\lambda_M^g = a_{Gg0} + a_{Gg1}$ with $g = 0, 1$. We can therefore write

$$r_{2g,2g+1} = \lambda_G \gamma G_g - \lambda_M^g \quad (5.45)$$
Recall also that the relevance relations in the rules for explicit similarity are expressed in equations (5.8) and (5.9). The exponents of the cross-terms are thus proportional to the difference in relevance between predicates of equal gender and predicates of different gender,

$$r_{2g,2g+1} = \gamma Gg N(\rho G - \rho Gg),$$  \hspace{1cm} \text{(5.46)}

where $\rho Gg$ and $\rho G$ are as indicated above. So the cross-terms in the prior probability over $B$ have non-zero exponents precisely if there are differences in the relevances between $Q$-predicates of identical and different gender.

It can be seen very easily that certain systems of rules for the underlying predicate families $G$ and $M$ are equivalent to a single $\lambda \gamma$ rule for $Q$-predicates. We only need to assume $a_{Gg} = a_{Gg0} + a_{Gg1}$, or in terms of the parameters in the system of rules

$$\lambda G \gamma Gg = \lambda_M g.$$  \hspace{1cm} \text{(5.47)}

This choice of parameters indeed reduces the system of rules to a single $\lambda \gamma$ rule. If we identify the numbers of observations $n_G = n_G$, $n_{Gg} = n_M^g$ and $n_{Mm} = n_{2g+m}$, the resulting system of rules is exactly identical to a single $\lambda \gamma$ rule with the parameters $\lambda = \lambda_G$ and $\gamma_{(2g+m)} = \gamma Gg \gamma_{Mm}$.

The Jacobian and virtual observations. Finally, it is illustrative to connect the Jacobian determinant to the system of prediction rules for underlying predicates, in particular to the notion of virtual observations. Consider a uniform prior probability over the space $B$, corresponding to the exponents $\lambda \gamma q = c_q = 1$ for all $q$. Sometimes these exponents are called the virtual observations of $Q^q$, since they are added to the number of actual observations $n_q$ in the prediction rules. Now if we translate the uniform prior over $B$ to a prior over $A$ directly, the resulting exponents $a_{Gg}$ and $a_{Ggm}$ are all 1, so that we obtain a uniform prior again. But because of the inverse Jacobian $J^{-1}(\alpha)$, the exponents $a_{Gg}$ are raised with 1, so that $a_{Gg} = 2$ and $a_{Ggm} = 1$, and correspondingly $p(H_\alpha) \sim \alpha_0 \alpha_1$. The fact that the prior over $A_G$ that corresponds to the uniform prior over $B$ is not flat is thus entirely due to the Jacobian deriving from the transformation between $B$ and $A_G$.

Now if we consider the exponents as resulting from virtual observations, the correction factor given by the Jacobian may be given a very natural interpretation: the exponents must be such that all combinations of predicates $G^g$ and $M^m$ have exactly one virtual observation. The thing to note is that, in terms of the underlying predicate family $G$, virtual observations for all $Q$-predicates entail two observations in each predicate $G^g$, so that indeed we must
have $a_{G,p} = 2$. More generally, we may think of the Jacobian as a function that supplements lost virtual observations after transformations of the hypotheses space. In other words, the number of virtual observations is the aspect of the prior probability that is supposed to remain intact during the transformation. This is very helpful in constructing Jacobians for more complicated parameter transformations than the one above.

5.5 GENERAL MODEL

It is not straightforward to construct the general model from the explicit analogy models. In the first subsection, some considerations will temper the ambition of finding a general model. The second subsection, however, will develop a tentative general model, but it will also reveals a problem for this model. The last subsection speculates on a specific solution with limited parameter freedom.

5.5.1 A more modest aim

Ansatz for the general model. The preceding discussion suggests that the inductive relevance between $Q^v$ and $Q^w$ may be modelled by multiplying the prior probability over the partition $B$ with a term $(\theta_v + \theta_w)^{r_{vw}}$. As indicated, the exponent $r_{vw}$ expresses the difference between two relevances: the relevance between $Q^v$ and $Q^w$ on the one hand, and the relevance between either one of these on the one hand, and predicates $Q^q$ with $q \neq v, w$ on the other. This suggests the following form for an overall analogy prior:

$$p(H_\theta) \sim \prod_{q<4} \theta_q^{c_q-1} \prod_{v<w<4} (\theta_v + \theta_w)^{r_{vw}}. \quad (5.48)$$

The predictions are determined entirely by the exponents $c_q$ and $r_{vw}$. Note that the prior can easily be generalised to settings with any number of $Q$-predicates.

In the Carnapian $\lambda \gamma$ rule we are able to connect the initial probabilities $\gamma_q$ for $q < 4$ and a learning rate $\lambda$ with the exponents of the prior probability over $B$. Now let us say that we are given a set of initial probabilities $\gamma_q$, a learning rate $\lambda$, and a vector of symmetric relevance relations, $\rho$. The challenge for the general model of analogical predictions then is to provide the exponents $c_q$ and $r_{vw}$ that correspond to these initial values. It can be noted immediately that there are an equal number of free components of $\gamma$, $\lambda$ and $\rho$, namely $3 + 1 + 6 = 10$, as there are free exponents in the above probability density. This suggests that there is indeed a unique solution for the representation problem. However, as will become clear, a complete representation is too much to ask for within the context of the present chapter.
Reasons for modesty. This section has a more modest aim: it presents a prior of the above form that on certain assumptions incorporates a vector of relevance relations and a learning rate, assuming a kind of initial symmetry between the $Q$-predicates. This modesty is motivated by a number of reasons. First, the procedure for incorporating relevance relations cannot be generalised in any straightforward way from the model for explicit similarity. We have to make assumptions to pin down the relations between the relevances and the prior probability. Second, it is very difficult to derive an analytic expression for the initial probabilities $\gamma_q$ from the analogy prior over $B$. Therefore we cannot directly control the initial probabilities implicit in the general model. Third, the learning rate $\lambda$ turns out to have a different role in the general model. And finally, the present model only deals with the relations between prior and relevance relations for the specific case of the party centre example. This will suggest some general guidelines for determining the exponents $c_q$ and $r_{vw}$ starting from a general relevance vector $\rho$, but full generality is not achieved.

5.5.2 Towards a general model

Encoding relevance relations. With this aim in mind, consider the relevance relations of the example on party centres. With $G$, $M$ and $W$ referring to underlying predicate families on gender, marital status and traditional wedding rings respectively, we can denote the components of the vector of equation (5.3) in the following way:

$$\rho = (\rho_{G0}, \rho_{G1}, \rho_{M0}, \rho_{M1}, \rho_{W0}, \rho_{W1}).$$

(5.49)

Here $\rho_{G0}$ refers to the components $\rho(2g, 2g + 1)$, since $Q^{2g}$ and $Q^{2g+1}$ have the underlying predicate $G^g$ in common. Other components of $\rho$ in equation (5.3) may be explicated in similar fashion: $\rho_{Mm}$ refers to $\rho(m, 2 + m)$ and $\rho_{Ww}$ to $\rho(1 - w, 2 + w)$.

Recall the basic pattern for relevance relations that derive from the partition $A_G$. The relevances $\rho_{G0}$ and $\rho_{G1}$ can be determined separately, while the other relevances, $\rho_{Mm}$ and $\rho_{Ww}$ for $m, w = 0, 1$, may be fixed on some average value. As before I denote this latter average value with $\rho_G$, while $\rho_G = (\rho_{G0} + \rho_{G1})/2$. Similar configurations of relevance relations may be derived from the partitions $A_M$ and $A_W$, enabling us to independently determine $\rho_{M0}$ and $\rho_{M1}$, or $\rho_{W0}$ and $\rho_{W1}$, and determine the average values $\rho_M$ and $\rho_W$. Now to incorporate the combination of these relevance relations into a single prior over $B$, we must imagine that the single prior results from the priors over the analogy partitions, which each incorporate specific aspects of the relevance vector. The exponents
\( c_q \) in the prior thus relate to the priors over all three analogy partitions. The exponents \( r_{vw} \), on the other hand, are related only to the priors over the analogy partitions associated with the combination of \( v \) and \( w \).

**Reducing the number of free parameters.** It seems natural to combine the three models for explicit similarity by multiplying the priors over \( \mathcal{B} \) corresponding to these models. This means that the exponents \( a_{Ggm}, a_{Mmw} \) and \( a_{Wwg} \) sum up to \( c_q \). The sum of the relevance vectors for the explicit similarity models may then serve as the relevance vector associated with the resulting prior. Alternatively, we may consider the product of the relevance vectors. However, all such straightforward combination procedures result in poor representations: different general analogy models are connected to the same analogical prior, and any analogical prior may be read as the result of a multitude of relevance vectors. We have too much freedom in choosing the exponents \( a_{Ggm}, a_{Mmw} \) and \( a_{Wwg} \) on the basis of the values for \( c_q \) and \( r_{vw} \).

The following employs a combination procedure in which the exponents \( c_q \) are not taken as built up from the exponents \( a_{Ggm}, a_{Mmw} \) and \( a_{Wwg} \) separately, but in which these latter exponents are each taken to be equal to the exponents \( c_q \). Every pair from the predicates \( G^g, M^m \) and \( W^w \) determines a unique predicate \( q \) in the family \( \mathcal{Q} \). With some algebra we can obtain the relations

\[
a_{Ggm} = c_{2g+m}, \quad a_{Mmw} = c_{2m-2w+4wm}, \quad a_{Wwg} = c_{1-w+2w}. \tag{5.50}
\]

So while we were considering 4 independent exponents in all three analogy partitions, these exponents must all coincide with the same set of 4 exponents \( c_q \). The priors over the analogy partitions that may be taken to underly the prior over \( \mathcal{B} \) are thus limited in a specific way: their exponents must conform to the corresponding values of the \( c_q \). The number of free exponents in the models for explicit similarity is thus reduced, so that the representation problems of the preceding paragraph are avoided.

**Combining the explicit similarity models.** Now consider the relation between the exponents in the analogy priors and the exponents \( r_{vw} \) in the prior over \( \mathcal{B} \). It can be noted that the values of \( a_{Gg}, a_{Mm} \) and \( a_{Ww} \) are implicit to the values of \( c_q \) and \( r_{vw} \) in an analogical prior over \( \mathcal{B} \), according to the relations (5.50) and the further relations

\[
\begin{align*}
  r_{2g,2g+1} &= a_{Gg} - a_{Gg0} - a_{Gg1}, \tag{5.51} \\
  r_{m,2+m} &= a_{Mm} - a_{Mm0} - a_{Mm1}, \tag{5.52} \\
  r_{1-w,2+w} &= a_{Ww} - a_{Ww0} - a_{Ww1}. \tag{5.53}
\end{align*}
\]
This means that we may encode the values of $a_G$, $a_M$ and $a_W$ in a prior over $B$ relative to a choice for the exponents $a_G$, $a_M$ and $a_W$.

The above equations specify how the exponents in the prior over $B$ relate to the exponents in the priors over the analogy partitions. We can now concentrate on the function of the exponents in the priors over the analogy partitions in expressing relevance relations. First consider $a_G$, $a_M$ and $a_W$. Their values are directly related to the average relevances $\bar{\rho}_G$, $\bar{\rho}_M$ and $\bar{\rho}_W$ according to

$$\bar{\rho}_G = a_{G0} + a_{G1},$$
$$\bar{\rho}_M = a_{M0} + a_{M1},$$
$$\bar{\rho}_W = a_{W0} + a_{W1}. (5.54)$$

Any vector of relevance relations $\rho$ thus determines the values of the sums on the right side of the equations (5.54). Moreover, by fixing these sums of exponents we also encode the given relevance vector in the prior up to averages for the pairs of relevance relations, because we can write

$$\rho_G = \rho_M + \rho_W - \rho_G,$$
$$\rho_M = \rho_W + \rho_G - \rho_M,$$
$$\rho_W = \rho_G + \rho_M - \rho_W. (5.55)$$

Thus the average size of the relevance among predicate pairs is implicit to the pairwise sums of the exponents $a_G$, $a_M$ and $a_W$.

**Overspecification.** Unfortunately, if we combine the priors of the analogy partitions into a single prior over $B$ we run into a problem. As indicated, the size of $\rho_G$ is encoded in the sizes of pairwise sums of exponents, to wit $a_{G0} + a_{G1}$, $a_{M0} + a_{M1}$ and $a_{W0} + a_{W1}$. But if we follow the relations that hold in the model for explicit similarity, the sizes of $\rho_{G0}$ and $\rho_{G1}$, and thereby of $\rho_G$, are also determined by equation (5.8). With equations (5.34) and (5.35) these are in turn determined by the exponents $c_q$, and the ratios $a_{G0}/(a_{G0}+a_{G1})$ and $a_{G0}/(a_{G0}+a_{G1})$. The thing to note is that these exponents and ratios are independent of the size of $a_{G0} + a_{G1}$, or of any other such pairwise sum. So the size of $\rho_G$ seems to be doubly encoded in the combined analogy prior over $B$: once by the pairwise sums, and once by the exponents and ratios. The same can be said of the average relevances $\rho_M$ and $\rho_W$. So the prior cannot encode all the relevances that we wish, since there are too few free parameters left once they are restricted to agree on the sizes of the average relevances.

In other words, restriction (5.50) leads us into a problem after all. The number of free exponents in the separate analogy models matches the number
of free exponents in the combined model, but somehow the restrictions chosen do not allow us to straightforwardly generalise the separate analogy models.

5.5.3 Tentative solution

Selective generalisation. The above combination procedure need not be considered as a complete failure. A more constructive reading is that the general model is not in all aspects a generalisation of the model for explicit similarity, and that only specific aspects of the explicit similarity model can be taken over into the general one. The following exposition runs along this line. It simply assumes that the pairwise sums of exponents fix the pairwise averages of the relevances according to equations (5.54) and (5.55). With this kept fixed, the analogy prior incorporates initial probabilities and learning rates as well as possible. After that, the remaining parameter freedom is used to pin down the differences between the pairs $\rho_{Gg}$, $\rho_{Mm}$, and $\rho_{Ww}$.

It must be stressed that this subsection is rather speculative. Within the limits set by the assumption of equations (5.54) and (5.55), it presents a model that is constructed by playing with the analogy priors and predictions in the program Mathematica™. Specifically, I have looked at the influence of varying exponents in the prior on the predictions for a number of numerical examples. The examples of the next section may give some justification for the eventual model, but I am convinced that a derivation of some general model is possible. I have unfortunately not been able to find it.

Constructing an analogy prior. As for the values of $\gamma_q$ and $\lambda$ in relation to the analogical prior over $\mathcal{B}$, I will make two simplifying assumptions. First, we may assume that the values of the exponents $a_{Gg}$, $a_{Mm}$ and $a_{Ww}$ are pairwise identical:

$$a_{G0} = a_{G1}, \quad a_{M0} = a_{M1}, \quad a_{W0} = a_{W1}.$$  \hspace{1cm} (5.56)

With equation (5.54) and assumption (5.56), the exponents $a_{Gg}$, $a_{Mm}$ and $a_{Ww}$ are determined completely. The idea behind the assumption is that we force the resulting initial predictions to be at least close to symmetric. Specifically, in all the priors over analogy partitions that may underly the prior over $\mathcal{B}$, the exponents that concern the leading predicate are equal. At the end of the next section I return to the approximated initial symmetry resulting from that assumption.
The second assumption concerns the learning rate $\lambda$. As in the Carnapian framework, it is given by the total number of virtual observations:

$$\sum_q c_q = \lambda. \quad (5.57)$$

In the analogy model, the value of $\lambda$ is related to the rate at which the size of the analogy effects diminish over time. But it does not straightforwardly select a specific learning rate. There is a sense in which the learning rate $\lambda$ is also an expression of relevance. The larger $\lambda$ is, the slower the prediction of an observation of $Q^w$ diminishes with the observation of $Q^v$. It therefore seems natural to choose

$$\lambda = \frac{\rho_G + \rho_M + \rho_W}{3}. \quad (5.58)$$

At the end of the next section I shall return to the exact function of $\lambda$ in the analogy prior, and consider variations on the value assumed here.

Certain aspects of the relevance relations remain to be captured in the prior probability assignment, namely the differences between the relevances within the pairs, such as between $\rho_{G0}$ and $\rho_{G1}$. Further, the only freedom left in the parameters of the prior is in the division of the total number of virtual instances $\lambda$ over the separate $c_q$. Note that we must fix three such differences, for which we have exactly three free parameters available. The above discussion makes clear that we cannot simply employ the relation (5.46) to fix the $\rho_{Gg}$ separately. However, we may be able to employ this relation to connect the difference between the relevances $\rho_{G0}$ and $\rho_{G1}$ to the exponents $c_q$. The idea here is that instead of carrying over the relation (5.46) into the general model, we may be able to carry over a weaker relation that can be derived from it.

Taking the difference between $\rho_{G0}$ and $\rho_{G1}$ as the example case, and using the fact that $a_{G0} = a_{G1}$, we can write

$$\rho_{G0} - \rho_{G1} = 2(r_{23} - r_{01})$$

$$= 2(c_2 + c_3 - c_0 - c_1), \quad (5.59)$$

which exactly meets this desideratum. In the same way we can write for the other two pairs of relevances

$$\rho_{M0} - \rho_{M1} = 2(c_1 + c_3 - c_0 - c_2), \quad (5.60)$$

$$\rho_{W0} - \rho_{W1} = 2(c_0 + c_3 - c_1 - c_2). \quad (5.61)$$

With these three relations, we have completely fixed the values for the exponents $c_q$. And because of the relations (5.51), (5.52) and (5.53), we thereby also fix the values for the exponents $r_{vw}$. 
**Overview of the model.** We have now completed the construction of the analogy prior on the basis of a vector of relevance relations, initial symmetry and a learning rate. Let me summarise the procedure. If we are given a vector of relevance relations \( \rho \), we first calculate the averages \( \rho_G, \rho_M \) and \( \rho_W \). With these averages, equations (5.55) and (5.54), and assumption (5.56) we can then determine the exponents \( a_G, a_M, a_W \). The assumption ensures approximate initial symmetry. With assumption (5.58) and equations (5.59) to (5.61) we can subsequently determine the values for the exponents \( c_q \). Here the assumption relates to the rate at which the analogy effects diminish. Finally, using equations (5.50) to (5.53) we may finally fix the exponents \( r_{vw} \).

Before investigating some properties of the proposed model, it is important to stress again that this model is in many ways incomplete. For one thing, I have not proposed any relation between relevance relations and analogy priors for cases in which the aforementioned assumptions are violated. Because of this, many analogy priors cannot be linked to a relevance vector. However, I must leave a more complete model to future research.

### 5.6 Qualitative and numerical characterisations

In this section I describe the general analogy model further. First I discuss the analogy priors by considering their form on an analogy partition, after which I can relate the priors to so-called hyper-Carnapian models of analogical reasoning. Finally I provide some numerical examples of analogical predictions.

#### 5.6.1 The form of the analogy priors

**Nonfactorisable priors.** It is instructive to consider the general analogy prior in its functional form over one of the analogy partitions, for example over \( A_G \). Recall that this partition falls into three orthogonal subpartitions: one concerning the predicate family \( G \), associated with the exponents \( a_G \), and two concerning the predicate family \( M \) conditional on \( G_0 \) and \( G_1 \), associated with the exponents \( a_{0m} \) and \( a_{1m} \) respectively. Analogy effects between the predicates \( Q^{2g} \) and \( Q^{2g+1} \) may then be captured by Dirichlet priors over the separate subpartitions: by choosing \( a_{G_0} + a_{G_1} \) larger or smaller than \( a_G \) we can make the relevance between \( Q^{2g} \) and \( Q^{2g+1} \) larger or smaller than the relevance between pairs of predicates that do not have the gender in common. These differences between the Dirichlet priors correspond with the terms \( (\theta_{2g} + \theta_{2g+1})^{r_{2g,2g+1}} \) in the general analogy prior. If the latter are the only analogical terms in the prior
over $B$, the prior can be factorised into separate functions over the orthogonal subpartitions of $A_G$, and thus can be dealt with completely independently.

Now imagine that, starting from a product of Dirichlet priors over $A_G$, we want to express additional relevance relations between predicates of different gender. Intuitively, we want the prior probability over $A_G$ to be such that, if we update the probability over the hypotheses concerning the family $M$ conditional on $G^0$, we implicitly update the probability over the hypotheses concerning the family $M$ conditional on $G^1$, and vice versa. In other words, we must relate the independent Dirichlet parts of the prior probability. The thing to note is that such relations are exactly realized by the additional terms in the prior over the analogy partition corresponding to the terms $(\theta_v + \theta_w)^{r_{vw}}$ with $v = 0, 1$ and $w = 2, 3$. As an example, consider a relevance relation between $Q_0$ and $Q_3$. This is associated with the term $(\theta_1 + \theta_3)^{r_{13}}$ in the general analogy prior. The translated term, $(\alpha_{G0}\alpha_{G01} + \alpha_{G1}\alpha_{G11})^{r_{13}}$, relates the priors over the subpartitions in the appropriate way. Note finally that because of such terms, the resulting prior cannot be factorised into separate functions over orthogonal subpartitions anymore.

**Hyper-Carnapian rules.** There is yet another way to illustrate the role of the analogical terms in the general model, which connects to the hyper-Carnapian analogical prediction rules of Skyrms (1993) and Festa (1997). Leaving aside the underlying relevance vector and the initial probabilities for the moment, consider the role of the analogy term in the following prior:

\[
p(\theta) \sim \theta_0\theta_1\theta_2\theta_3(\theta_2 + \theta_3) \\
= \theta_0\theta_1\theta_2^2\theta_3 + \theta_0\theta_1\theta_2\theta_3^2.
\] (5.62)

This prior consists of two parts, which can each be associated with a $\lambda\gamma$ prediction rule $pr_{\lambda\gamma}(n_{Qq}, n_Q)$ of equation (5.18). Both these rules have $\lambda = 9$, but the initial probabilities $\gamma_q$ vary. For the first term we have $\gamma_2 = 1/3$ while $\gamma_q = 2/9$ for $q \neq 2$, while the second term entails $\gamma'_2 = 1/3$ while $\gamma'_q = 2/9$ for $q \neq 3$. The predictions resulting from the above prior are therefore a mixture of two $\lambda\gamma$ rules, one using $\gamma_q$ and one $\gamma'_q$.

Such mixtures are called hyper-Carnapian prediction rules. For the predictions generated by the above hyper-Carnapian rule we can write

\[
p(Q_{n+1}|E_n) = p(H_{\lambda\gamma}|E_n)pr_{\lambda\gamma}(n_{Qq}, n_Q) + p(H_{\lambda\gamma'}|E_n)pr_{\lambda\gamma'}(n_{Qq}, n_Q),
\] (5.63)

where the hypotheses $H_{\lambda\gamma}$ have likelihoods given by the rules $pr_{\lambda\gamma}$. It can now be seen that the hyper-Carnapian rules capture analogical considerations. On
the observation $Q^3$, the rule is adapted in two ways: first the two $\lambda \gamma$ rules are adapted to enhance the probability for future observations of $Q^3$, but the update over the hypotheses also enhances the probability of the rule for which $Q^3$ has a higher initial probability. This latter update does not affect the probabilities for future observations of $Q^0$ or $Q^1$, but it does lower the probability for $Q^2$ to raise that of $Q^3$. This is exactly the kind of analogy effect that may be expected from the term $(\theta_2 + \theta_3)$ in an analogy prior.

### 5.6.2 Numerical Examples

Predictions deriving from the general analogy prior cannot usually be written down in any simple analytic form. While positive exponents $r_{vw}$ can still be captured in terms of extended hyper-Carnapian rules, negative analogy exponents seem to make analytic expressions impossible. The remainder of this section is concerned with two examples of general analogical predictions, making use of the numerical integration module of Mathematica\textsuperscript{TM}. I present the examples to show that the predictions generated by the general model agree with the predictions that can be expected on the basis of the relevance relations. For the sake of easy calculations I focus on examples with natural numbers.

**Symmetric relevance.** The first of these examples is the simpler one, as it has some inherent symmetries. Consider the following relevance vector:

$$\rho = \langle \rho_G^0, \rho_G^1, \rho_M^0, \rho_M^1, \rho_W^0, \rho_W^1 \rangle = \langle 28, 8, 22, 22, 14, 14 \rangle.$$  \hfill (5.64)

To calculate the exponents of the analogy partition, note that $\rho_G = 18$, $\rho_M = 16$ and $\rho_W = 20$, so that $a_G = 9$, $a_M = 8$ and $a_W = 10$. Furthermore, note that $\sum_q c_q = 18$, and $c_0 + c_2 - c_1 - c_3 = c_1 + c_2 - c_0 - c_3 = 0$, while $c_0 + c_1 - c_2 - c_3 = 10$. It follows that $c_0 = c_1 = 7$ and $c_2 = c_3 = 2$, and with that it follows that $r_{10} = -5$, $r_{23} = 5$, $r_{02} = r_{13} = -1$ and $r_{12} = r_{03} = 1$. The analogy prior is thus completely specified. This prior leads to the predictions given in the table. To illustrate the analogy effects, the table shows the initial probabilities, and the predictions after 10 observations of the same predicate $Q^q$ for each $q$. I abbreviate $E_{10}^q = Q_1^q \cap Q_2^q \cap \ldots \cap Q_{10}^q$. 

...
A number of things may be remarked on these results. First, the initial probabilities are only approximately symmetric. Rounded off they are the same, but they differ from each other at the fifth decimal. The analogy terms associated with \( G \), to wit \((\theta_{2g} + \theta_{2g+1})^{r_{2g,2g+1}}\), normally cancel the differences between the exponents \(c_{2g} + c_{2g+1}\) for \(g = 0, 1\). But the analogy terms related to \( M \) and \( W \) interfere with these terms and cause minor imbalances. Second, it is notable that the predictions respect the ordering of the relevance relations. Moreover, the predicates \( Q^0 \) and \( Q^1 \) gain less probability from their own occurrences than the predicates \( Q^2 \) and \( Q^3 \), which is in line with the fact that the latter are on average less relevant to other predicates than the former. Unfortunately, the differences between the predictions are not in any way linear in the differences in relevance. Third, there is perfect symmetry between predicates of equal gender, \( Q^{2g} \) and \( Q^{2g+1} \). This can be seen from the fact that the prior probability is invariant under permutation of these predicates. It is therefore not surprising that the predictions of the first and the second two lines are identical up to this permutation.

No symmetries. The second example breaks with this symmetry. Consider the following vector of relevance relations:

\[
\rho = (28, 8, 20, 24, 16, 12).
\] (5.65)

This example also has \( \rho_G = 18 \), \( \rho_m = 16 \) and \( \rho_W = 20 \), so that again \( a_G = 9 \), \( a_m = 8 \), \( a_W = 10 \) and \( \sum_q c_q = 18 \). But the example further has \( c_0 + c_2 - c_0 - c_2 = 2 \), \( c_1 + c_2 - c_0 - c_3 = 2 \), while again \( c_0 + c_1 - c_2 - c_3 = 10 \). It follows that \( c_0 = 6 \), \( c_1 = 8 \) and \( c_2 = c_3 = 2 \), and with that it follows that \( r_{01} = -5 \), \( r_{23} = 5 \), \( r_{02} = 0 \), \( r_{13} = -2 \) and \( r_{12} = 2 \), \( r_{03} = 0 \). This completely specifies the analogy prior, and we may again consider the predictions after \( E_{10}^q \).
5.7 Other models of analogical predictions

Let me briefly compare the resulting models for analogical predictions with some other models in the literature. First I relate the present model to a number of alternative prediction rules from Carnap-Hintikka inductive logic. After that I consider the hyper-Carnapian systems by Skyrms and Festa, and finally I turn to the model of Maher.

Carnap-Hintikka inductive logic. The general aim in Carnap-Hintikka inductive logic is to derive a class of prediction rules from a number of natural assumptions or principles. One of these principles then is an expression of analogy by similarity, other principles may be regularity, exchangeability, the convergence to relative frequencies, initial symmetry with respect to predicates, positive probability for confirmed universal generalisations, and instantial relevance. Combinations of these principles are employed in the derivation of classes of analogical prediction rules.

<table>
<thead>
<tr>
<th>Predicate q</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(Q^q_1)$</td>
<td>0.252</td>
<td>0.248</td>
<td>0.250</td>
<td>0.250</td>
</tr>
<tr>
<td>$\rho(0, q)$</td>
<td>-</td>
<td>28</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>$p(Q^0_1</td>
<td>E^0_{10})$</td>
<td>0.499</td>
<td>0.193</td>
<td>0.169</td>
</tr>
<tr>
<td>$\rho(1, q)$</td>
<td>28</td>
<td>-</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>$p(Q^1_1</td>
<td>E^1_{10})$</td>
<td>0.199</td>
<td>0.466</td>
<td>0.154</td>
</tr>
<tr>
<td>$\rho(2, q)$</td>
<td>20</td>
<td>16</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>$p(Q^2_1</td>
<td>E^2_{10})$</td>
<td>0.168</td>
<td>0.154</td>
<td>0.582</td>
</tr>
<tr>
<td>$\rho(3, q)$</td>
<td>12</td>
<td>24</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>$p(Q^3_1</td>
<td>E^3_{10})$</td>
<td>0.147</td>
<td>0.172</td>
<td>0.097</td>
</tr>
</tbody>
</table>

More or less the same remarks can be made on these results. The main thing is that the predictions still respect the ordering in the relevance relations, and that the probability that predicates gain from their own occurrence is still dependent on the average relevances. Note further that the symmetry in the initial probabilities is more disturbed than in the first example. Finally, note that the symmetry between predicates of equal gender is broken. With this relevance vector, $Q^0$ is on average less relevant to the other predicates than $Q^1$. In line with this, the predicate $Q^0$ gains more probability from its own occurrence than $Q^1$. This effect is almost absent for $Q^2$ and $Q^3$, but again the model is not exactly correct.
It is instructive to position the present models in terms of these principles. First, the above model trivially satisfies regularity: none of the finite sequences of observations are deemed impossible from the onset. Second, the predictions deriving from the general model are by definition exchangeable, and therefore show convergence to eventual relative frequencies in the sequence of observations, if there are any. This can be seen from the fact that the models are defined by probability assignments over the partition $\mathcal{B}$, and by the standard convergence results derived for this partition. Third, initial symmetry of predicates can in principle be obtained by choosing the analogy prior over $\mathcal{B}$ appropriately. But as we have seen in the foregoing, encoding both a vector of relevance relations and initial symmetry in the analogy prior is not straightforward. Finally, the above model does not give positive probability to universal generalisations. Hypotheses $H_\alpha$ have infinitesimal probability, since the probability is always distributed over a continuum of hypotheses $\mathcal{A}$. The infinitesimal probability is thus also assigned to those $H_\alpha$ in which one or more components of $\alpha$ are extremal. However, there is a rather natural extension of the above schemes in which these sets are given strictly positive measure. Universal generalisations are certainly not excluded by the above models.

The principle of instantial relevance must be given separate attention. It is noteworthy that instantial relevance need not always be satisfied for predictions resulting from a prior over $\mathcal{B}$. This principle may be violated exactly because the prior probability over the hypotheses space $\mathcal{A}$ need not be factorisable into independent marginals. In terms of the example, we may consider the presence of wives much more probable than that of maidens if there are very few women in the party centre, while we may consider the presence of maidens much more probable than that of wives in the case that there are very few men. Now, observing a wife in the party centre has a combined effect: first of all it makes the presence of women more probable, and within the group of women it shifts the probability from maidens to wives. However, it may happen so that the former effect is much stronger than the second: after a single woman we hardly expect any further men. But because maidens are considered much more probable than wives if there are hardly any men, the probability of maidens may eventually benefit more from observing a wife than the probability of wives itself. It is not difficult to construct the prior that encodes the above effects numerically. However, I have not been able to check whether there are such priors in the class of analogy priors defined above. Moreover, because predictions defined on $\mathcal{B}$ always converge to the correct relative frequencies, the effect sketched above is necessarily a short term one.
Finally, consider the relation between the analogy model and Carnap-Hintikka inductive logic more generally. Recall that a special case of the analogy model is presented by a system of $\lambda\gamma$ rules that models explicit similarity effects. In turn, these systems of prediction rules have the single Carnapian $\lambda\gamma$ rule as special case. The model may therefore be considered as an extension of Carnap-Hintikka inductive logic. However, in contrast to the direct prediction rules of this logic, the predictions of the general analogy model can only be arrived at by numerical approximation of an integration over statistical hypotheses. Furthermore, while the model consists of a class of analogy priors, there is no claim that this class is somehow the definitive explication of analogical reasoning. In these ways the general analogy model falls outside Carnap-Hintikka logic.

Skyrms and Festa. Let me now turn to the models for analogy reasoning by Skyrms and Festa, which employ hyper-Carnapian prediction rules. The idea of such rules originates from Skyrms, but Festa explores the rules further to define a proper class of analogical prediction rules. The standard illustration involves a wheel of fortune with four equally large segments, labelled with the four quarters of the compass. At every turn in a direction chosen at random, the chance of stopping at some segment is unknown but constant. It is further given that the axis of the wheel is slightly eccentric. Finding the wheel in some segment will favour this segment in next predictions, but it will on the whole also favour the two neighbouring segments in comparison to the opposing segment. The hyper-Carnapian prediction rule proposed for this is a mixture of four $\lambda\gamma$ rules. The rules have equal $\lambda$s, and for each of them the $\gamma$s of the segments are chosen as $\frac{1}{2}$ for the segment favoured by the bias, $\frac{1}{5}$ for the neighbouring segments, and $\frac{1}{10}$ for the opposing one. The rules differ in that each of them has its own favoured segment. As it turns out, the resulting predictions then show similarity effects between all pairs of neighbouring segments. That is, if we find an instance of north, east and west are favoured more than south, and so on.

The hyper-Carnapian models are similar to, but also different from the present analogy models. They provide alternative ways for defining analogy priors over the partition $\mathcal{B}$, and in the specific case in which the analogy terms have positive exponents, the present analogy model is also a kind of hyper-Carnapian model. However, this correspondence fails for priors with negative analogy exponents $r_{vw}$. Furthermore, I see a difficulty in making sense of the prior probability proposed by hyper-Carnapian rules, which is related to the difficulties noted in Maher (2000, 2001). On the partition $\mathcal{B}$, hyper-Carnapian
rules are defined as mixtures of Dirichlet priors. After an observation we must update every Dirichlet distribution separately, and apart from that we must adapt the weights assigned to these different distributions according to the predictions that the separate distributions generate. But assigning probabilities to Dirichlet priors seems rather unnatural. Such probabilities cannot be interpreted as probabilities over hypotheses, but must really be seen as probabilities assigned to different priors over the hypotheses, that is, as a kind of second-order probability. And if we can also define analogy priors by means of a single probability function over one space of hypotheses, introducing such higher order probabilities seems a high price to pay.

The model of Maher. Finally, some attention must be given to the model for analogical predictions proposed by Maher. This model is again similar to the present model in important respects. It generates predictions on $Q$-predicates by defining an analogy prior over the partition $B$, and moreover, it employs underlying predicates such as $G$ and $M$ in the definition of this prior. A drawback is that the model of Maher is limited to two underlying predicate families. Maher uses these families for defining a set of hypotheses within the partition $B$ for which the families are statistically independent. While a Dirichlet distribution over $B$ assigns zero probability to this set, Maher assigns a positive probability to it. He shows that conditional on the independence, the predictions on $Q$-predicates can be represented as a product of $\lambda \gamma$ rules for the underlying predicates. He further derives a single $\lambda \gamma$ rule conditional on the dependence of the underlying predicates. The prediction rule generated by the combined prior over $B$ is a mixture of this single $\lambda \gamma$ rule and the product of the two $\lambda \gamma$ rules for the underlying predicates.

As said, the model of this chapter has a lot in common with this model. However, the model of Maher does not employ the possibilities with underlying predicates completely. It uses predictions concerning such predicates on the condition that they are independent. By contrast, the model for explicit analogy also employs hypotheses concerning underlying predicates on the condition that these predicates are statistically dependent. The present model generalises this to incorporate dependencies between all three predicate families that may underlie the four $Q$-predicates. Moreover, there are no principal problems with defining analogy priors for predictions on larger numbers of $Q$-predicates.

Advantages of the present model. Let me emphasise some of the advantages of the models presented in this chapter, when compared with other models in the literature. First, the models of this chapter show analogical predictions as
5.8. CONCLUSION

The result of a Bayesian scheme using hypotheses. As argued in the first part of this thesis, the Bayesian scheme ensures that the predictions are valid, and clearly reveals the inductive assumptions underlying these predictions. Second, and as shown in the preceding chapter, the present models allow for generalisations to more predicate families. Third, and as elaborated in this chapter, the models provide access to the inductive relevance vector that is inherent to a prior probability assignment over the algebra. This latter feature is also present in Kuipers (1984), but both Festa (1997) and Maher (2000) fail to provide any such connection between analogical prediction rules and relevances.

5.8 Conclusion

Summary and moral. This chapter presents a Bayesian model for exchangeable analogical predictions. First relevance relations were characterised, and related to the models for explicit similarity of the preceding chapter. Then the chapter presented a scheme that employs hypotheses on $Q$-predicates for generating predictions. Exchangeable predictions for $Q$-predicates were represented with a prior over the partition $B$. It was shown how the $Q$-predicates may be translated in combinations of underlying predicates $G$, $M$ and $W$. The partitions of hypotheses concerning these predicates, denoted $A_G$, $A_M$ and $A_W$, were seen to be equivalent to the original partition $B$. Dirichlet priors over the separate parts of these partitions capture the exchangeable analogical predictions for explicit similarity. Transforming these priors back to the partition $B$ suggested a form for a general analogy prior over this latter partition. Finally, the chapter proposed a relation between relevance relations and this prior on the basis of the relevance relations expressed in the explicit similarity model.

In closing, let me draw a general moral from the above models. It is that some of the problems in defining analogical predictions within Carnap-Hintikka inductive logic may be solved by shifting perspective twice. The first shift concerns the use of statistical hypotheses, and the explicit use of Bayes’ rule in accommodating observations. That is, we represent prediction rules with a Bayesian update over hypotheses. In this perspective, exchangeable rules can be characterised with a prior over the partition $B$. But the priors that lead to analogical predictions are hard to define in the parameter space associated with $B$. In the second shift, this difficulty is solved by transforming the partition $B$ into the analogical partitions, which have a differently structured parameter space but are otherwise equivalent. These parameter spaces supply the conceptual means for defining priors that lead to analogical predictions.
Bayesian logic and the Carnapian programme. Let me concentrate on the first perspectival shift. It concerns the logical framework of analogical predictions. It illustrates that Carnap-Hintikka inductive logic can be viewed as part of a wider logic of inductive Bayesian inference, as developed in chapters 1 to 3. In this logic of inductive inference, validity is determined exclusively by the probability axioms, which here include Bayesian conditioning. Further, the inductive relevance of observations for each other is not inherent to the choice of language and the assumption of some further principles. Instead the inductive relevance is inherent to a partitioning of the observation algebra into statistical hypotheses and a prior probability assignment over this partition. If we are given the observational algebra, we have complete freedom in choosing these inductive assumptions. I believe that both the expression of inductive assumptions in a partition of hypotheses and the neutral way of incorporating observations into the inductive methods present valuable conceptual advantages.

While this offers a useful perspective on analogical predictions, I am aware that it also masks one of the intentions of Carnap-Hintikka inductive logic. This intention is to present a normative theory of inductive predictions from first principles. If this intention is applied to the problem of analogical predictions, it is to provide a class of rules resulting in predictions that conform to a certain characterisation of analogy, and that are logically valid. By contrast, this chapter only presents some examples of Bayesian models for analogical predictions. For those who do not share the intentions of Carnap-Hintikka inductive logic, the examples may already suffice: they are exemplars for models for analogical predictions. But from the standpoint of Carnap-Hintikka inductive logic itself, the above examples can perhaps best be taken as providing a framework and some starting points for a further normative discussion.

Using transformations between partitions. Concentrating on the second shift, note first that the use of underlying predicates follows quite naturally from the model for explicit similarity, as given in the preceding chapter. It is perhaps hard to come up with a transformation of the partition $\mathcal{B}$ into, for example, $\mathcal{A}_G$ if there is no independent reason for thinking of the partition $\mathcal{A}_G$ in the first place. On the other hand, it is strictly speaking inessential what the partition on underlying predicate families refers to. The transformation procedure from the predicate $Q$ to underlying predicates can simply be taken as a formal tool for expressing relations between the $Q$-predicates. Similarly, we do not need a natural and independent description of resulting $Q$-predicates in order to employ explicit similarity relations in predictions over underlying predicates.
Another aspect of the second perspectival shift is more important for the general line of this thesis. In chapter 3 I have argued that the Bayesian scheme offers a better control over the inductive assumptions inherent to inductive predictions, by linking these assumptions to the choice of a partition. But this chapter and the preceding one reveal another advantage of the Bayesian scheme. It is that the scheme, once the partition has been chosen, allows us to express further inductive assumptions in a specific prior probability function over the partition. And for this we are free to transform the partition in such a way that the function becomes more easily accessible. Thus, not only the choice of a partition is a tool for making inductive assumptions, the partition itself is also a tool in defining a prior over the partition, which may express further inductive assumptions. I refer to chapter 9 for a further discussion of this idea.