# 17 The GCM model of categorization

WITH RUUD WETZELS

## 17.1 The GCM model

The Generalized Context Model (GCM: Nosofsky, 1984, 1986) is an influential and empirically successful model of categorization. It is intended to explain how people make categorization decisions in a task where stimuli are presented, one at a time, over a sequence of trials, and must be classified into one of a small number of categories (usually two) based on corrective feedback.

The GCM assumes that stimuli are stored as exemplars, using their values along underlying stimulus dimensions, which correspond to points in a multidimensional psychological space. The GCM then assumes people make similarity comparisons between the current stimulus and the exemplars, and base their decision on the overall similarities to each category.

A key theoretical component of the GCM involves selective attention. The basic idea is that, to learn a category structure, people selectively attend to those dimensions of the stimuli that are relevant to distinguishing the categories. Nosofsky (1984) showed that selective attention could help explain previously puzzling empirical regularities in the ease with which people learn different category structures (Shepard, Hovland, & Jenkins, 1961).

We consider category learning data from the "Condensation B" condition reported by Kruschke (1993).[1] This condition is shown in Figure 17.1, and involves eight stimuli—consisting of line drawings of boxes with different heights, with an interior line in different positions—divided into two groups of four, to make Category A and Category B stimuli. Kruschke (1993) collected data from 40 participants over 8 consecutive blocks of trials, within which each stimulus was presented once in a random order. These data can be summarized by $y_{ik}$, the number of times the $i$th stimulus was categorized as belonging to Category A by the $k$th participant, out of the $t = 8$ trials on which it was presented. In an analysis that does not consider individual differences, the data can be further summarized as $y_i = \sum_k y_{ik}$, the total number of times all participants categorized the $i$th stimulus into Category A, out of $t = 40 \times 8$ total presentations.

[1] The Kruschke (1993) category learning experiments involved corrective feedback after every trial. Usually, the GCM is not applied to this sort of task, but to categorization decisions made without feedback after a training period involving feedback.

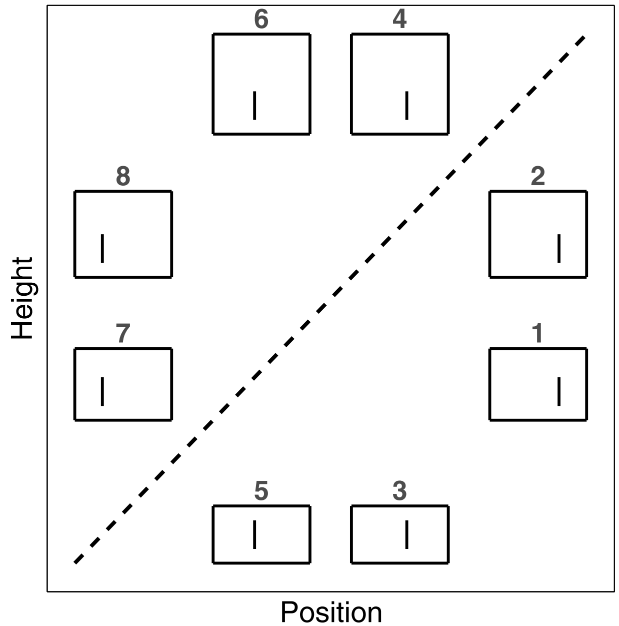
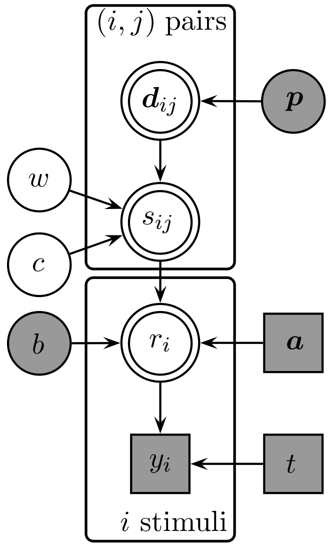



**Fig. 17.1** Condensation category structure "B" from Kruschke (1993). Stimuli above the diagonal belong to category B and stimuli below the diagonal belong to category A.

A graphical model representation of the GCM, as applied to the group data, is shown in Figure 17.2. There are two stimulus dimensions, so the $i$th stimulus is represented by the point $(p_{i1}, p_{i2})$. The first dimension has attention weight $w$ with $0 \leq w_d \leq 1$, and the second dimension then has attention weight $(1 - w)$. These weights act to "stretch" attended dimensions, and "shrink" unattended ones, so that the psychological distance between the $i$th and $j$th stimuli is $d_{ij} = w\,|p_{i1} - p_{j1}| + (1-w)\,|p_{i2} - p_{j2}|$. The similarity between the $i$th and $j$th stimuli is $s_{ij} = \exp(-cd_{ij})$, where $c$ is a generalization parameter. The overall similarity of the $i$th stimulus, when it is presented, to Category A is $s_{iA} = \sum_{j \in A} s_{ij}$. The graphical model uses these in an unbiased choice rule, so that the probability that the $i$th stimulus will be classified as belonging to Category A, rather than Category B, is $r_i = bs_{iA} / \left(bs_{iA} + (1-b)\,s_{iB}\right)$, with $b = 0.5$. The observed decisions themselves are then given by $y_i \sim \text{Binomial}\,(r_i, t)$.

The script `GCM_1.txt` implements the graphical model in WinBUGS. Note that it collects posterior predictive samples for the decision data in the variable `predy`:

```
# Generalized Context Model
model{
  # Decision Data
  for (i in 1:nstim){
    y[i] ~ dbin(r[i],t)
    predy[i] ~ dbin(r[i],t)
  }
```




$$
\begin{aligned}
c &\sim \text{Uniform}(0,5)\\
w &\sim \text{Uniform}(0,1)\\
b &\leftarrow \tfrac{1}{2}\\
d_{ij}^{m} &\leftarrow |p_{im} - p_{jm}|\\
s_{ij} &\leftarrow \exp\left\{-c\left(wd_{ij}^{1} + (1-w)\,d_{ij}^{2}\right)\right\}\\
r_i &\leftarrow \frac{b\sum_j a_j s_{ij}}{b\sum_j a_j s_{ij} + (1-b)\sum_j (1-a_j) s_{ij}}\\
y_i &\sim \text{Binomial}(r_i, t)
\end{aligned}
$$

Fig. 17.2 Graphical model implementation of the GCM.

```
  # Decision Probabilities
  for (i in 1:nstim){
    r[i] <- sum(numerator[i,])/sum(denominator[i,])
    for (j in 1:nstim){
      tmp1[i,j,1] <- b*s[i,j]
      tmp1[i,j,2] <- 0
      tmp2[i,j,1] <- 0
      tmp2[i,j,2] <- (1-b)*s[i,j]
      numerator[i,j] <- tmp1[i,j,a[j]]
      denominator[i,j] <- tmp1[i,j,a[j]] + tmp2[i,j,a[j]]
    }
  }
  # Similarities
  for (i in 1:nstim){
    for (j in 1:nstim){
      s[i,j] <- exp(-c*(w*d1[i,j]+(1-w)*d2[i,j]))
    }
  }
  # Priors
  c  ~ dunif(0,5)
  w  ~ dbeta(1,1)
  b <- 0.5
}
```

The code GCM_1.m or GCM_1.R applies the model to the Kruschke (1993) data. The joint posterior distribution of the generalization parameter $c$ and the attention weight parameter $w$ is shown as a scatter-plot in Figure 17.3. The key result, in terms of the theory of category learning and selective attention, is that the attention parameter $w$ lies between about 0.5 and 0.7. This can be interpreted as showing that

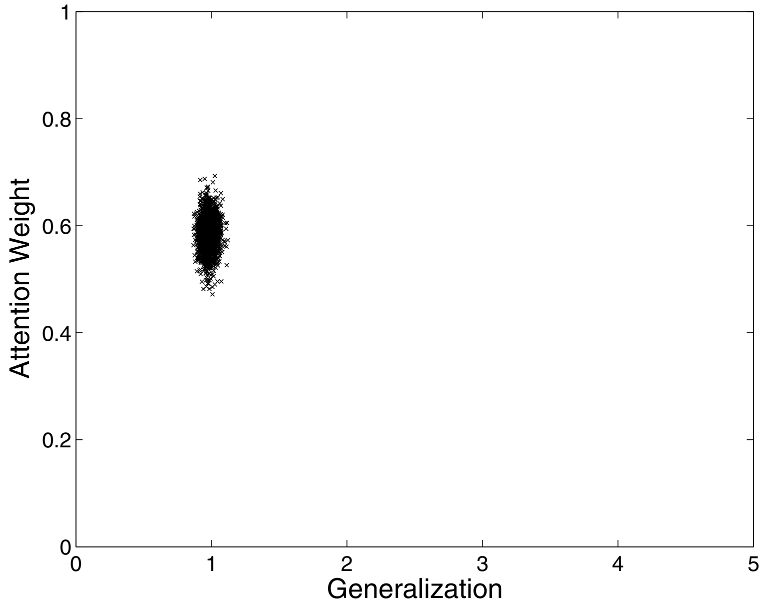



Fig. 17.3 Joint posterior distribution over attention $w$ and generalization $c$ parameters of the GCM, when applied to the Kruschke condensation data.

people give significant attention to both dimensions, although they are probably focusing a little more on the line position than the rectangle height.

This is consistent with the condensation task design of the category structure. It is clear from Figure 17.1 that both dimensions are relevant in determining to which category a stimulus belongs, and so the shared attention result makes sense.

## Exercises

**Exercise 17.1.1** Setting $b = 0.5$ to make the decision rule unbiased seems reasonable, since there are two alternatives with equal numbers of equally-often-presented stimuli. But, the assumption can be easily examined. Change the model so that the bias parameter $b$ is given a uniform prior over the range 0 to 1, and is inferred from the data. Summarize the findings from this model, and compare them to the results from the original model.

**Exercise 17.1.2** Figure 17.4 shows a posterior predictive analysis of the modeling. The average $y_i$ counts are shown for each of the 8 stimuli, overlaid on gray violin plots showing the posterior predictive distributions. Also shown, by the broken lines, are individual participant data. These are linearly scaled from the individual count of 8, to the group count of 320, to allow visual comparison of the number of times. From this figure, what do you conclude about the ability of the GCM to describe the group data? What do you conclude about the adequacy of the group data as a summary of human performance?

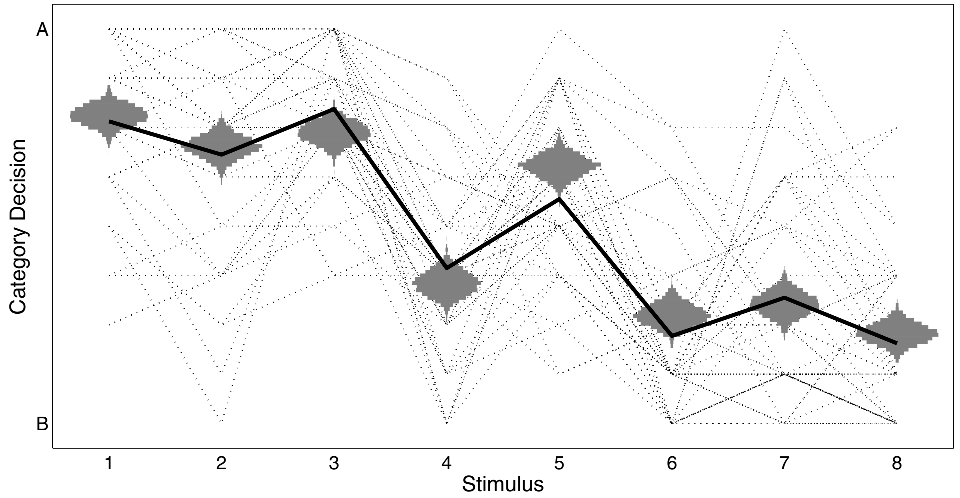



Fig. 17.4 Posterior predictive analysis of the GCM applied to the condensation data.

## 17.2 Individual differences in the GCM

Figure 17.4 suggests that there are significant individual differences in the categorization data. Figure 17.5 shows the individual data more clearly, with each subject's decisions about each stimulus presented in a separate panel.

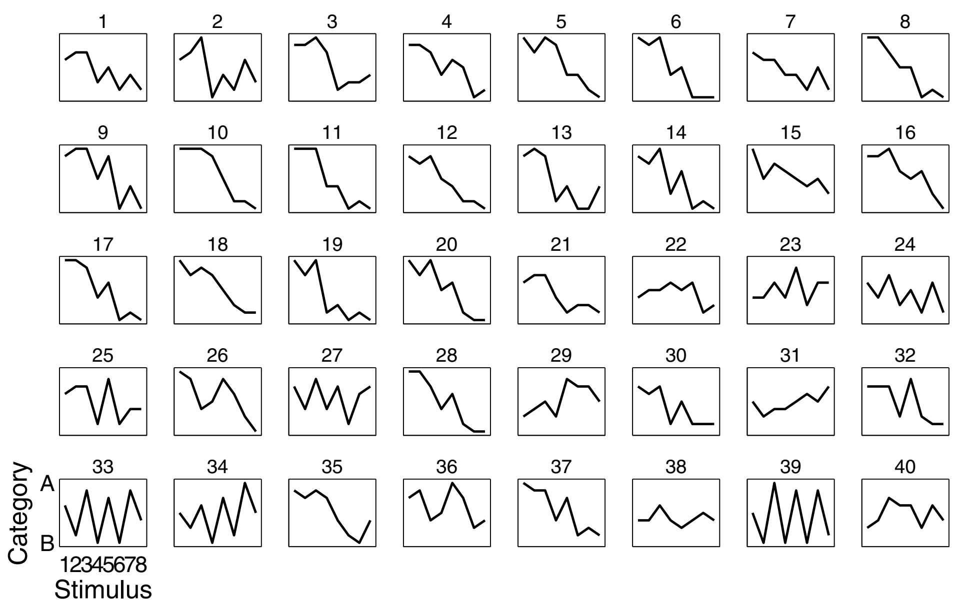
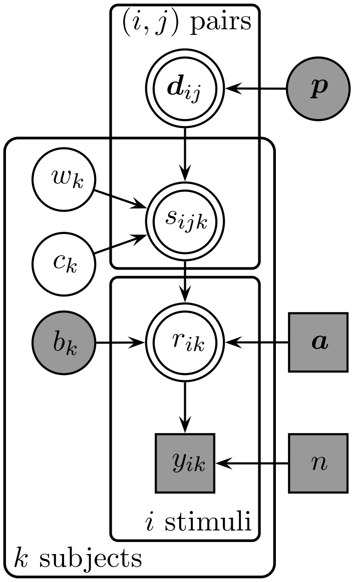



Fig. 17.5 Category decision data for all 40 subjects.




$$
\begin{aligned}
c_k &\sim \text{Uniform}(0,5)\\
w_k &\sim \text{Uniform}(0,1)\\
b_k &\leftarrow \tfrac{1}{2}\\
d_{ij}^m &\leftarrow |p_{im} - p_{jm}|\\
s_{ijk} &\leftarrow \exp\left\{-c_k\left(w_k d_{ij}^1 + (1-w_k)\, d_{ij}^2\right)\right\}\\
r_{ik} &\leftarrow \frac{b_k \sum_j a_j s_{ijk}}{b_k \sum_j a_j s_{ijk} + (1-b_k)\sum_j (1-a_j) s_{ijk}}\\
y_{ik} &\sim \text{Binomial}(r_{ik}, n)
\end{aligned}
$$

**Fig. 17.6** Graphical model implementation of the GCM allowing for full individual differences in attention and generalization.

The simplest way to start to investigate these differences using the GCM is to infer parameter values for each subject independently. A graphical model to do this is presented in Figure 17.6. It simply allows the $k$th subject to have their own generalization and attention parameters $c_k$ and $w_k$, and adds a plate replicating those parts of the GCM that depend upon the parameters. Notice that the extension to individual differences means the individual subject data, $y_{ik}$, for the $i$th stimulus and $k$th subject are now being modeled, rather than the summary data $y_i$.

The script `GCM_2.txt` implements the graphical model in WinBUGS:

```
# Generalized Context Model With Individual Differences
model{
  # Decision Data
  for (i in 1:nstim){
    for (k in 1:nsubj){
      y[i,k] ~ dbin(r[i,k],n)
      predy[i,k] ~ dbin(r[i,k],n)
    }
  }
  # Decision Probabilities
  for (i in 1:nstim){
    for (k in 1:nsubj){
      r[i,k] <- sum(numerator[i,k,])/sum(denominator[i,k,])
    }
  }
  # Base Decision Probabilities
  for (i in 1:nstim){
    for (j in 1:nstim){
```

```
      for (k in 1:nsubj){
        numerator[i,k,j] <- equals(a[j],1)*b*s[i,k,j]
        denominator[i,k,j] <- equals(a[j],1)*b*s[i,k,j]
                            + equals(a[j],2)*(1-b)*s[i,k,j]
      }
    }
  }
  # Similarities
  for (i in 1:nstim){
    for (j in 1:nstim){
      for (k in 1:nsubj){
        s[i,k,j] <- exp(-c[k]*(w[k]*d1[i,j]+(1-w[k])*d2[i,j]))
      }
    }
  }
  # Parameters and Priors
  for (k in 1:nsubj){
    c[k] ~ dunif(0,5)
    w[k] ~ dbeta(1,1)
  }
  b <- 0.5
}
```

The code GCM_2.m or GCM_2.R applies the model to the Kruschke (1993) data. The joint posterior distributions of the generalization parameter $c$ and the attention weight parameter $w$, for all 40 subjects, are represented in Figure 17.7. The filled circles show the posterior mean for each individual subject, and the lines radiating from these points connect the mean to a small number of randomly selected samples of the joint posterior for that subject.

### Exercises

**Exercise 17.2.1** Compare the inferences about the attention parameter based on an individual subject analysis in Figure 17.7 with those based on a no individual differences analysis in Figure 17.3. Give a psychological interpretation of these differences, in terms of how the subjects selectively attended to the stimulus dimensions.

**Exercise 17.2.2** Three of the individual subject joint posterior distributions—for Subjects 3, 31, and 33—in Figure 17.7 are labeled. These three subjects lie in different areas of the parameter space, and so possibly correspond to different sorts of categorization behavior. Look at the individual data in Figure 17.5 for these three subjects, and give a short description of the differences in their categorization decisions.

## 17.3 Latent groups in the GCM

The individual differences displayed in terms of basic behavior in Figure 17.5 and through inference about GCM parameters in Figure 17.7 provide information that

 https://doi.org/10.1017/CBO9781139087759.019

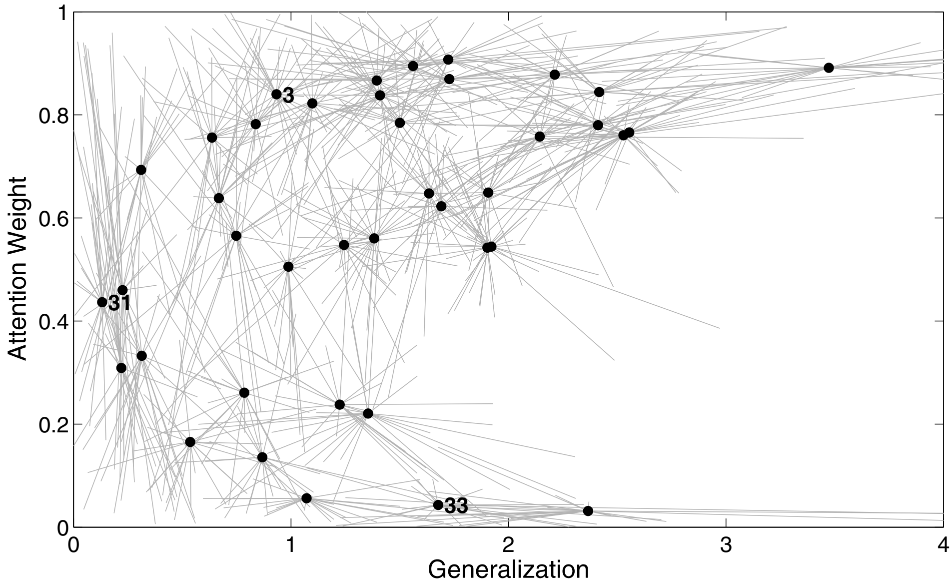
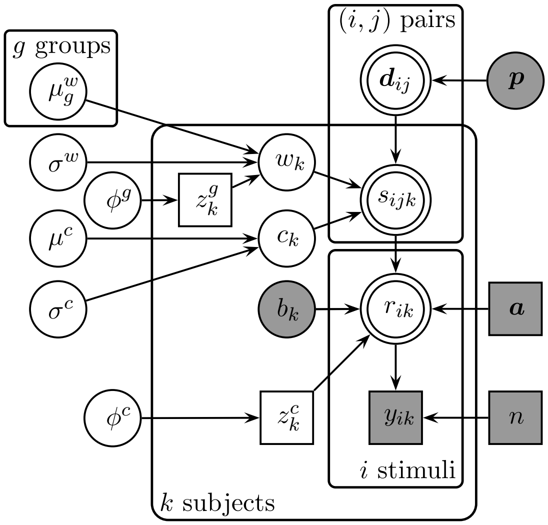



**Fig. 17.7** Posterior distributions for each subject, for the attention $w$ and generalization $c$ parameters, inferred by the GCM with full individual differences.

can help formulate more constrained models of the variation in category learning in this data set. It is not the case that there is one correct model of these differences but, rather, that some models will be able to describe the variability better than others, and provide more useful insights into the underlying psychology.

One worthwhile model assumes there are three groups of subjects, and is shown as a graphical model in Figure 17.8. This model is based on the insight that there are three types of behavior, well exemplified by Subjects 3, 31, and 33. In the first group, subjects like Subject 31 seem to categorize each stimulus roughly equally often in each category, consistent with guessing behavior. These subjects might be interpreted as being contaminants, who are not attempting to do the task diligently. These subjects have GCM parameter posterior distributions with low generalization values, and large uncertainty about their attention values, as seen in Figure 17.7. This makes sense, since low values of the generalization parameter will lead to high similarity between all stimuli, and near-equal summed similarity for both category responses in this task. A much simpler way to account for the behavior of subjects like Subject 31, however, is not to use the GCM to account for their behavior at all, but to use a separate contaminant model that just sets their response probabilities to be $r_{ik} = \frac{1}{2}$ for all stimuli.

The groups of subjects represented by Subjects 3 and 33, meanwhile, show more thoughtful categorization behavior, but have different patterns of responding to some stimuli. Recall from Figure 17.1 that stimuli 1, 2, 3, and 5 belong to category A, and stimuli 4, 6, 7, and 8 belong to category B. Subjects like Subject 3 make




$$
\begin{aligned}
\mu_1^w, \delta &\sim \text{Uniform}(0,1)\\
\mu_2^w &\leftarrow \min\left(1, \mu_1^w + \delta\right)\\
\sigma^w &\sim \text{Uniform}(0.01,1)\\
\mu^c &\sim \text{Uniform}(0,5)\\
\sigma^c &\sim \text{Uniform}(0.01,3)\\
\phi^c, \phi^g &\sim \text{Uniform}(0,1)\\
z_k^c &\sim \text{Bernoulli}(\phi^c)\\
z_k^g &\sim \text{Bernoulli}(\phi^g)\\
c_k &\sim \text{Gaussian}(\mu_c, \tfrac{1}{(\sigma^c)^2})_{\mathcal{I}(0,5)}\\
w_k &\sim \begin{cases} \text{Gaussian}(\mu_1^w, \frac{1}{(\sigma^w)^2})_{\mathcal{I}(0,1)}, \\ \ \text{if } z_k^c = 0,\ z_k^w = 0 \\ \text{Gaussian}(\mu_2^w, \frac{1}{(\sigma^w)^2})_{\mathcal{I}(0,1)}, \\ \ \text{if } z_k^c = 0,\ z_k^w = 1 \end{cases}\\
b_k &\leftarrow \tfrac{1}{2}\\
d_{ij}^m &\leftarrow |p_{im} - p_{jm}|\\
s_{ijk} &\leftarrow \exp\left\{-c_k\left(w_k d_{ij}^1 + (1 - w_k)\, d_{ij}^2\right)\right\}\\
r_{ik} &\leftarrow \begin{cases} \frac{b_k \sum_j a_j s_{ijk}}{b_k \sum_j a_j s_{ijk} + (1-b_k) \sum_j (1-a_j) s_{ijk}}, \\ \ \text{if } z_k^c = 0 \\ \frac{1}{2}, \ \text{if } z_k^c = 1 \end{cases}\\
y_{ik} &\sim \text{Binomial}(r_{ik}, n)
\end{aligned}
$$

**Fig. 17.8** Graphical model implementation of the GCM allowing for three latent groups. One group of subjects are contaminants, the other two groups use the GCM, but with different selective attention.

almost all of their errors on stimuli 4 and 5. This is consistent with a focus on the position dimension, corresponding to large values of the $w$ attention parameter, which would assign these two stimuli incorrectly. Subjects like Subject 33, however, categorize stimuli 2 and 7 poorly. This is consistent with a focus on the height dimension, corresponding to low values of the attention parameter $w$.

The script `GCM_3.txt` implements the graphical model in WinBUGS. Note that posterior predictive distributions are generated at the group level, and that a three-valued classification variable `z` is constructed showing the group membership of each subject:

```
# Generalized Context Model With Contaminants and Two Attention Groups
model{
  # Decision Data
  for (i in 1:nstim){
    # Subjects
    for (k in 1:nsubj){
      y[i,k] ~ dbin(r[i,k],n)
    }
    # Groups
    for (g in 1:3){
      predyg[g,i] ~ dbin(rpredg[g,i],n)
    }
  }
  # Decision Probabilities
  for (i in 1:nstim){
    for (k in 1:nsubj){
```

```
      r[i,k] <- equals(zc[k],0)*sum(numerator[i,k,])/sum(denominator[i,k,])
                + equals(zc[k],1)*0.5
    }
    for (g in 1:2){
      rpredg[g,i] <- sum(numeratorpredg[g,i,])/sum(denominatorpredg[g,i,])
    }
    rpredg[3,i] <- 0.5
  }
  # Base Decision Probabilities
  for (i in 1:nstim){
    for (j in 1:nstim){
      for (k in 1:nsubj){
        numerator[i,k,j] <- equals(a[j],1)*b*s[i,k,j]
        denominator[i,k,j] <- equals(a[j],1)*b*s[i,k,j]
                              + equals(a[j],2)*(1-b)*s[i,k,j]
      }
      for (g in 1:2){
        numeratorpredg[g,i,j] <- equals(a[j],1)*b*spredg[g,i,j]
        denominatorpredg[g,i,j] <- equals(a[j],1)*b*spredg[g,i,j]
                                   + equals(a[j],2)*(1-b)*spredg[g,i,j]
      }
    }
  }
  # Similarities
  for (i in 1:nstim){
    for (j in 1:nstim) {
      for (k in 1:nsubj){
        s[i,k,j] <- exp(-c[k]*(w[k]*d1[i,j]+(1-w[k])*d2[i,j]))
      }
      for (g in 1:2){
        spredg[g,i,j] <- exp(-cpredg[g]*(wpredg[g]*d1[i,j]+(1-wpredg[g])*d2[i,j]))
      }
    }
  }
  # Subject Parameters
  for (k in 1:nsubj){
    c[k] ~ dnorm(muc,lambdac)I(0,)
    w[k] ~ dnorm(muw[zg1[k]],lambdaw)I(0,1)
  }
  # Predicted Group Parameters
  for (g in 1:2){
    wpredg[g] ~ dnorm(muw[g],lambdaw)I(0,1)
    cpredg[g] ~ dnorm(muc,lambdac)I(0,)
  }
  # Priors
  b <- 0.5
  # Latent Mixture
  phic ~ dbeta(1,1)
  phig ~ dbeta(1,1)
  for (k in 1:nsubj){
    zc[k] ~ dbern(phic)
    zg[k] ~ dbern(phig)
    zg1[k] <- zg[k]+1
    z[k] <- equals(zc[k],0)*zg1[k] + 3*equals(zc[k],1)
  }
  # Mean Generalization
```

**Box 17.1 Enduring knowledge**

"[T]he most substantial and enduring advances [in cognitive science] have not been in the accumulation of empirical facts or the construction of models, but in the production of fruitful interactions between models and experimental research. Most experimental facts require continual reinterpretation and most models drop by the wayside like autumn leaves, but the results of interactions between models and experiments constitute most of our generalizable knowledge." (Estes, 2002, p. 3)

```
  muctmp ~ dbeta(1,1)
  muc <- 5*muctmp
  # Mean Attention
  muwtmp ~ dbeta(1,1)
  muw[1] <- muwtmp
  delta ~ dbeta(1,1)
  muw[2] <- min(1,delta+muw[1])
  # Standard Deviation Generalization
  sigmactmp ~ dbeta(1,1)
  sigmac <- max(.01,3*sigmactmp)
  # Standard Deviation Attention
  sigmawtmp ~ dbeta(1,1)
  sigmaw <- max(.01,sigmawtmp)
  # Precision
  lambdac <- 1/pow(sigmac,2)
  lambdaw <- 1/pow(sigmaw,2)
}
```

The code `GCM_3.m` or `GCM_3.R` applies the model to the Kruschke (1993) data, and generates two graphs that analyze the key results. The first of these is presented in Figure 17.9, and shows the latent assignment probabilities of each subject into the three possible groups. It is clear that almost all of the subjects are confidently classified into one of the three groups. This is a strong indication that the groups proposed by the model are useful characterizations of the data at the level of individual subjects. If a single group was not able to account well for individual data, the natural outcome of Bayesian inference is model averaging, in which a blend of different groups is used to describe the data, and there is significant posterior mass for more than one assignment.

Figure 17.10 shows a posterior predictive analysis for each of the three groups. Each panel corresponds to a group, with the squares showing the predictive distribution of the model. The thick lines show the average categorization behavior for subjects classified as belonging to the group, as determined from the mode of their posterior group membership shown in Figure 17.9. The thin lines show the individual behavior of these subjects. For each of the three groups, the qualitatively different pattern of categorization behavior seems well described by the posterior predictive distribution.

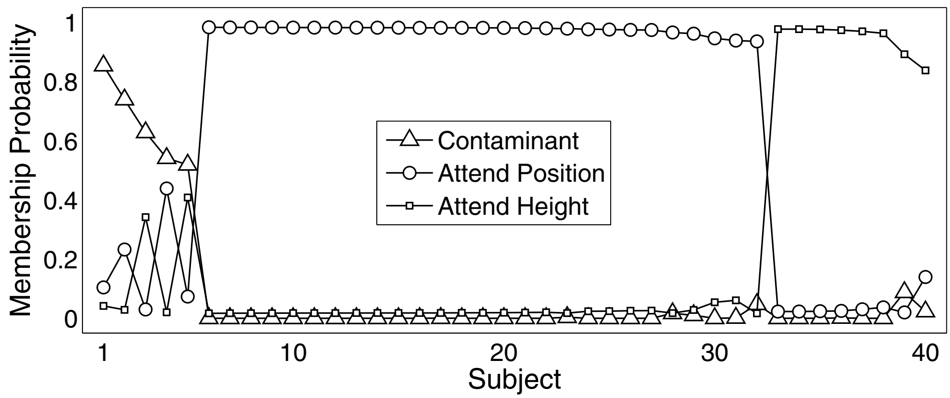



Fig. 17.9 The probability of assignment of each subject. Note that subjects have been ordered to make the information easy to parse, and this ordering is different from the subject ordering in the raw data, as shown in Figure 17.5.

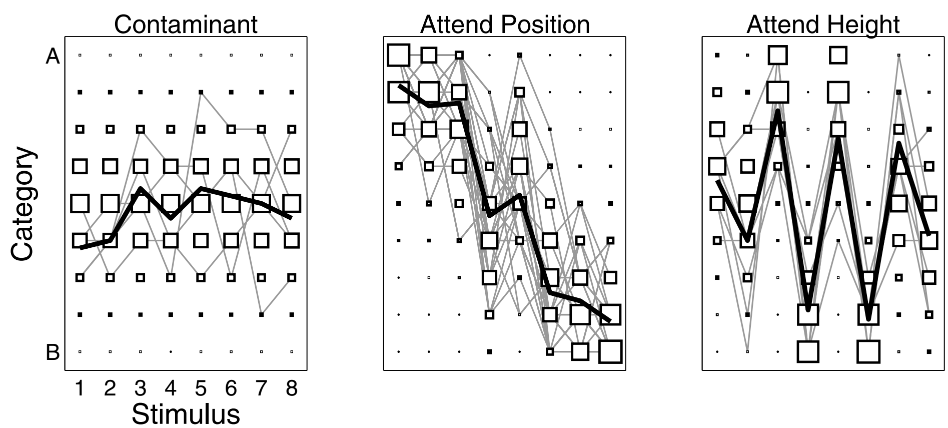



Fig. 17.10 Posterior predictive distributions for the latent, attend position and attend height groups, shown by squares, and the average and individual category decisions for subjects assigned to each group, shown by thick and thin lines, respectively.

## Exercises

**Exercise 17.3.1** The analyses presented focus on the inferred group membership, and the posterior predictive distributions for each group. These might be two of the most important inferences, but they are not the only available or useful ones. Extend the analysis by considering the posterior distributions of the inferred proportion of contaminant subjects $\phi^c$, and the difference in group mean attention $\delta$, giving psychological interpretations for both.

**Exercise 17.3.2** Construct the posterior distribution for the probability that a subject is in the attend position group rather than the attend height group. This is not simply the posterior for the $\phi^g$ parameter.