Revista Colombiana de Estadística
Current Topics in Statistical Graphics
Diciembre 2014, volumen 37, no. 2, pp. 319 a 339
DOI: http://dx.doi.org/10.15446/rce.v37n2spe.47940

Hierarchical Graphical Bayesian Models in Psychology

Modelos Bayesianos gráficos jerárquicos en psicología

Guillermo Campitelli[1,a], Guillermo Macbeth[2,b]

[1]School of Psychology and Social Science, Edith Cowan University, Mount Lawley, Australia

[2]Facultad de Ciencias de la Educación, Universidad Nacional de Entre Ríos, Argentina

Abstract

The improvement of graphical methods in psychological research can promote their use and a better comprehension of their expressive power. The application of hierarchical Bayesian graphical models has recently become more frequent in psychological research. The aim of this contribution is to introduce suggestions for the improvement of hierarchical Bayesian graphical models in psychology. This novel set of suggestions stems from the description and comparison between two main approaches concerned with the use of plate notation and distribution pictograms. It is concluded that the combination of relevant aspects of both models might improve the use of powerful hierarchical Bayesian graphical models in psychology.

***Key words*:** Visual Statistics, Graphical Models, Bayesian Statistics, Hierarchical Models, Psychology, Statistical Cognition.

Resumen

El mejoramiento de los métodos gráficos en la investigación en psicología puede promover su uso y una mejor compresión de su poder de expresión. La aplicación de modelos Bayesianos gráficos jerárquicos se ha vuelto más frecuente en la investigación en psicología. El objetivo de este trabajo es introducir sugerencias para el mejoramiento de los modelos Bayesianos gráficos jerárquicos en psicología. Este conjunto de sugerencias se apoya en la descripción y comparación entre los dos enfoques principales con el uso de notación y pictogramas de distribución. Se concluye que la combinación de los aspectos relevantes de ambos puede mejorar el uso de los modelos Bayesianos gráficos jerárquicos en psicología

***Palabras clave*:** cognición estadística, estadística Bayesiana, estadística visual, modelos gráficos, modelos jerárquicos, psicología.

[a]Lecturer. E-mail: gjcampitelli@gmail.com
[b]Professor. E-mail: g.macbeth@conicet.gov.ar

1. Introduction

Graphical models are powerful formal tools to represent dependencies and independence among variables, to make probabilistic inferences of conditional probabilities, and to estimate the values of parameters (Jordan 2004, Koller, Friedman, Getoor & Taskar 2007, Pearl 2009). This article focuses on the use of graphical models as tools to specify or illustrate hierarchical Bayesian models in psychology. More specifically, we concentrate on the visual features and the pieces of information represented in graphical hierarchical Bayesian models in psychology. Issues related to the formalism of the graphical model representation, algorithms to infer conditional probabilities, and machine learning procedures to estimate parameters are beyond the scope of this article (see for an exhaustive treatment of these issues Koller & Friedman 2009). Research into the role of graphs on the understanding of statistical information has shown that diagrams facilitate different kinds of reasoning (Bauer & Johnson-Laird 1993, Stenning & Oberlander 1995). It also emphasised that, in order to make research interesting for readers, graphs that represent complex information should be kept as simple as possible (Gray & Wegner 2013).

This article discusses two types of graphical models that have been used in psychology to represent hierarchical Bayesian models–graphical models with plate notation (Buntine 1994, Gilks, Thomas & Spiegelhalter 1994, Lee 2008) and graphical models with distribution pictograms (Kruschke 2010*a*, Kruschke 2010*b*). Moreover, it introduces a new type of graphical model that aims at combining the positive aspects of those two alternatives, while keeping the representation as simple as possible. In line with Gray and Wegner's proposal we aimed at generating an attractive graphical representation, for example, by adding 3D shapes to the graph. As indicated by a reviewer and by Wickens, Merwin & Lin (1994) 3D graphs sometimes improve understanding, but sometimes decrease understanding of statistical information. Therefore, we were very cautious when using 3D shapes. The article continues as follows. We first very briefly describe hierarchical Bayesian models, and then we discuss the use of graphical models to represent such models. After that we present the rationale of graphical models with plate notation, and we describe the introduction of hierarchical Bayesian graphical models in psychology. We then explain and discuss the use of graphical models with plate notation and graphical models with distribution pictograms. Lastly, we present the new type of representation, and we discuss advantages and disadvantages.

2. Hierarchical Bayesian Models

Bayesian analysis was introduced to psychology by Edwards, Lindman & Savage (1963). In this article we concentrate on the hierarchical aspect of the hierarchical Bayesian models (for an introduction to Bayesian statistics see, for example, Bolstad, 2007). Good (1980) indicated that there are three types of hierarchies that could be represented with hierarchical Bayesian models. One type is hierarchies about physical probabilities. For example, in a population of animals, there

is a probability that an animal belongs to a category of animal (e.g., mammal), a probability that it belongs to a species (e.g., dog), a probability that belongs to an age category (e.g., 1 year of age), and so forth.

The second type of hierarchy arises from the fact that subjective probabilities about events cannot be sharp (e.g., the probability of an article being accepted being 0.65474). One way of dealing with this is to express the confidence of this subjective probability, and represent it as a probability distribution of a higher type. The third type of hierarchy is a combination of the first two types. Lee (2008) takes a pragmatic approach, and considers that a Bayesian model is hierarchical when it is more complex than a Bayesian model with a set of parameters θ generating a set of data d through a likelihood function $f(.)$. Similarly, Kruschke's (2010*a*) chapter on hierarchical Bayesian models includes hyper-parameters, which are parameters that do not generate data, instead they affect other parameters.

3. Graphical Model Representation of Hierarchical Bayesian Models

Jordan (2004) indicated that a graphical model is a family of probability distributions represented in a directed or undirected graph. Graphs contain nodes –which represent random variables– connected by edges, which could be either directed (i.e., arrows) or undirected (i.e., lines). The most popular graphical models are Markov networks and Bayesian networks (or "Bayes nets"). In Markov networks all the edges are undirected whereas in the Bayesian networks all the edges are directed and the graph is acyclic. A graph is cyclic if it contains a cyclic path. A path is a sequence of nodes in which each node, except the first node in the sequence, receives a directed edge from the precedent node in the sequence (e.g., o→ o→ o→ o). When the last node in the sequence is the same as the first node, the path is cyclic, otherwise it is acyclic. In this article we focus on directed acyclic graphs (DAG). Although graphical models could be used both by frequentist and Bayesian approaches, they are very powerful tools to formulate complex models of joint probability distributions; hence, they have been very popular within the Bayesian approach (Jordan 2004). We present more details of graphical models in the following sections.

4. Plate Notation in Graphical Model Representations of Hierarchical Bayesian Models

The plate notation was introduced by Buntine (1994)[1] in order to solve the problem that graphical models were not representing the fact that the learning algorithms to estimate parameters had to do so over a repeated set of measured

[1]Note that Buntine indicated that the notion of plate was personally communicated to him by Spiegelhalter in 1993, but the similar notion of "replicated node" was developed by him independently, and that in order to maintain the same language he used the term plate.

variables. Figure 1 is an adaptation of Buntine's Figure 13, which shows how repetition of variables were represented without plates (panel a) and the same model represented with plates (panel b). The graphical model represents a biased coin toss experiment in which the probability of heads is θ , and the coin is tossed N times. Each time the coin is tossed the experimenter records whether it was a head or not. The hierarchical aspect of the model is apparent by the introduction of a prior distribution in the graph–that is, a Beta distribution with parameters $a = 1.5$ and $b = 1.5$–from which the parameter θ is generated.

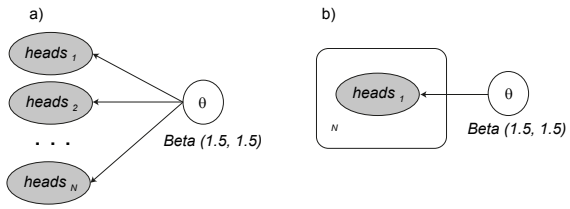

FIGURE 1: Adaptation of the plate graphical model (in panel b) presented by Buntine (1994) to represent the graphical model without plates in panel a. Instead of representing a number of nodes of the variable heads as in panel a, the plate representation surrounds one heads node by a plate, indicating that the model structure is repeated N times. Shaded nodes indicate observed variables and unshaded nodes indicate unobserved variables. The arrows show the dependencies between the variables.

There is nothing wrong with the model without plates, but when the number of variables and parameters increases the graphical model becomes difficult to represent in a limited space. The plate indicates that the variables within that plate are repeated the number of times indicated in the plate (in the model presented in Figure 1b, N times). This version of plate notation includes other important features: shaded nodes indicate observed data or variables with known values whereas unshaded nodes are unobserved variables. In other graphical models in the article Buntine (1994) used a double border to represent deterministic variables (the same notation was adopted by Lee (2008); see Figure 5.1, variables h_i and f_i). A note of interest is that Buntine included in the graph the Beta distribution from which the parameter θ is generated, but not the Bernoulli distribution by which each value in heads is obtained.

Figure 2 presents an adaptation of Gilks et al.'s (1994) graphical model with plates in their presentation of the BUGS (Bayesian inference Using Gibbs Sampling) language and sampler. (For historical reasons, it is probably important to note that the term "plate" was not used in this article). The model represents two research studies with the purpose of estimating disease risk associated to exposure to chemical agents: in the "disease study" the variables *jobtype* and *disease* were collected, and in the "exposure study" the variables *jobtype* and *exposuretoagents* were collected in another set of individuals. The model is the following:

$$D_i \sim Bernoulli(\theta_i), \tag{1}$$

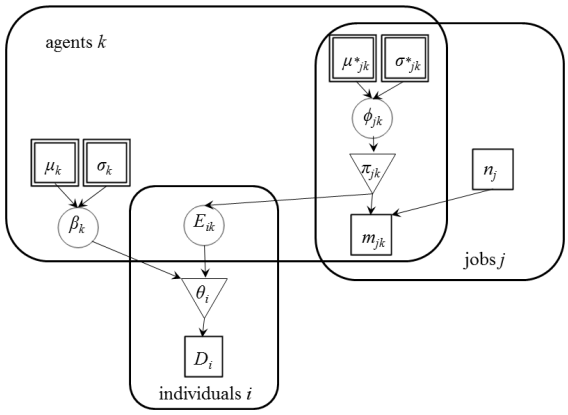

FIGURE 2: Graphical model with plates presented by Gilks et al. (1994). Circle nodes denote unobserved stochastic variables, square nodes with a single border represent observed stochastic variables (i.e., data), squares with double border denote fixed quantities in prior distributions, and triangles denote deterministic variables. The arrows represent the dependencies between variables. In this model all the arrows represent stochastic dependencies.

where D_i is the disease status of the ith individual and θ_i is his/her probability of disease

$$logit(\theta_i) = \beta_0 + \sum_{k=1}^{nA} \beta_k E_{ik}, \tag{2}$$

where E_{ik} is the unobserved exposure status (0 = unexposed; 1 = exposed) of the ith individual to the kth chemical agent, and $nA = 4$ is the number of chemical agents. The model for the exposure study is:

$$m_{jk} \sim Binomial(\pi_{jk}, n_j), \tag{3}$$

where $j = 1$ to nJ, $nJ = 2$ is the number of job types, π_{jk} is an exposure probability, n_j is the number of individuals in the exposure study who were in job type j, and m_{jk} is the number of those individuals in job type j who were exposed to chemical agent k. The connection between the disease study and the exposure study is indicated by:

$$E_{ik} \sim Bernoulli(\pi_{j(i)k}), \tag{4}$$

where $j(i)$ is the job type of individual i in the "disease study". The priors are denoted by:

$$\beta_k \sim Normal(\mu_k, \sigma_k^2) \tag{5}$$

and

$$logit(\pi_{jk}) = \phi_{jk}, \tag{6}$$

with

$$\phi_{jk} \sim Normal(\mu_k^*, \sigma_k^{*2}) \tag{7}$$

The graphical representation of this Bayesian hierarchical model in Figure 2 contains the plates i, j, k. The variables within each plate are replicated for each unit indicated in the plate. For example, D_i indicates that there is a value of disease for each individual i, and n_j that there is a value for the variable n for each job type j (e.g., 10 individuals are drivers, 25 individuals are teachers, etc.). The number of replications of variables that are contained in two plates (e.g., i and j) is the product of the number of units indicated in each plate (e.g., number of i units x number of k units). The replication aspect of this graphical model is identical to that of Buntine (1994).

However, there are a number of differences in the graphical representations. In Gilks et al. (1994) there are no shaded nodes, and the difference between observed and unobserved variables is given by the shape of the nodes. Circles denote unobserved stochastic variables, square nodes with a single border denote observed stochastic variables (i.e., data), squares with double border denote fixed quantities in prior distributions, and triangles denote deterministic variables. Another difference is that in Gilks et al. (1994) the name of the distributions are not included, and instead of including the total number of units per plate they present the unit index (i.e., i, j, k), the name of the units (i.e., individuals, jobs, agents), and the indices of the variables are located within the nodes.

Lunn, Thomas, Best & Spiegelhalter (2000) presented WinBUGS, which uses a graphical interface called DoodleBUGS. The graphical representation code in DoodleBUGS differs from the previous two versions. Figure 3 shows the simple example presented by Lunn et al. (2000), in which the graphical model represents a linear regression model expressed by

$$y_i \sim N(\mu_i, \tau^{-1}), \tag{8}$$

where y_i are observations measured at an experiment design points' x_i, $i = 1, \ldots, N, \tau$ is the inverse of the residual variance, and

$$\mu_i = \alpha + \beta x_i, \tag{9}$$

for $i = 1, \ldots, N$. α is the intercept parameter, β is the slope parameter, and both are unknown.

In this representation rectangular nodes denote known constants, and round nodes denote deterministic relationships or stochastic quantities. Stochastic dependence is represented by single-edged arrows and deterministic dependence is denoted by double-edged arrows. The plate includes both the index of the units (i.e., i) and the fact that the repetition is from i unit 1 to i unit N.

Summing up, a number of visual components of the graphical models have been used to represent hierarchical Bayesian models: shape of nodes, shade of nodes, type of arrows, indication of repetition, indices and inclusion of the distribution name. In the next section we describe the most popular plate notation that has been used to represent hierarchical Bayesian models in psychology and a new type of graphical model, which does not contain plates.

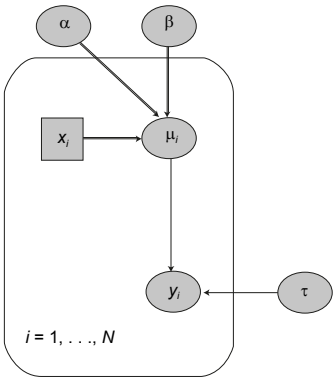

FIGURE 3: Graphical model with plates presented by Lunn et al. (2000). Rectangular nodes denote known constants, round nodes denote deterministic relationships or stochastic quantities. Stochastic dependence is represented by single-edged arrows and deterministic dependence is denoted by double-edged arrows.

5. Graphical Model Representation of Hierarchical Bayesian Models in Psychology

5.1. Lee's (2008) Hierarchical Bayesian Graphical Models with Plate Notation

Lee (2011) indicated that Bayesian statistics have been used in psychology in three ways. The first use of Bayesian statistics is to conduct data analysis. Edwards et al. (1963) introduced Bayesian analysis to psychology, and more recently a number of researchers (Dienes 2011, Kruschke 2010*a*, Kruschke 2010*b*, Lee & Wagenmakers 2005, Wagenmakers 2007) have advocated that Bayesian statistics should replace the Null Hypothesis Statistical Testing (NHST) paradigm, which is still the most popular paradigm in psychology. Another use of a Bayesian approach in psychology is to produce psychological models of how humans make inferences (see Griffiths, Kemp & Tenenbaum, 2008 for a detailed treatment of this approach). The third use –the one we focus on in this article–is to relate psychological models to data. To our knowledge Lee (2008) presented the first graphical model with plates to represent hierarchical Bayesian models in psychology. Lee "translated" three influential cognitive models in psychology –the multidimensional scaling (MSD) representation of stimulus similarity (Shepard, 1962, 1980), the generalized context model (GCM) account of category learning (Nosofsky, 1984, 1986), and a signal detection theory (SDT) account of reasoning (Heit & Rotello 2005)–into hierarchical Bayesian models. (see Figure 4).

The plate represents repetitions over N participants in a reasoning experiment (see 2005 Heit & Rotello for details of the experiment). Lee uses shaded nodes for observed variables and nodes without shading for unobserved variables. Continuous variables are represented by circles and discrete variables with squares; and

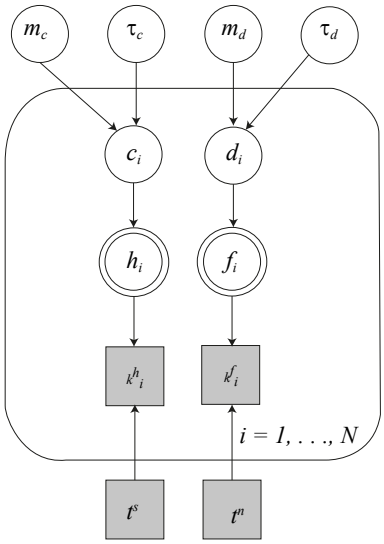

FIGURE 4: Graphical model with plate notation adapted from Lee (2008, Figure 10). Squares represent discrete variables, circles continuous variables, shaded nodes denote observed variables and unshaded nodes represent unobserved variables. The plate represents repetition of the model structure over units $i = 1$ to N. The arrows represent dependencies between variables.

stochastic and deterministic unobserved variables are denoted by single border and double border, respectively. In the model, d_i and c_i denote the ith individual's discriminability and bias, respectively. They are generated from a normal distribution with mean parameters m_d and m_c, respectively, and precision (i.e., 1/variance) parameters τ_d and τ_c; that is:

$$d_i \sim Normal(m_d, \tau_d) \tag{10}$$

$$c_i \sim Normal(m_c, \tau_c) \tag{11}$$

The priors of the mean and precision parameters are not shown in the graphical model, and their mean and precision have standard near non-informative priors $[m_d \sim Normal(0, \varepsilon), m_c \sim Normal(0, \varepsilon), \tau_d \sim Gamma(\varepsilon, \varepsilon), \tau_d \sim Gamma(\varepsilon, \varepsilon)$; where $\varepsilon = .001]$. The hit and false alarm rates are deterministic variables calculated by:

$$h_i = \Phi(0.5d_i - c_i) \tag{12}$$

$$f_i = \Phi(0.5d_i - c_i) \tag{13}$$

where $\Phi(.)$ is the standard cumulative Gaussian function. The count on hit rate and false alarms comes from a Binomial distribution with probabilities h_i and f_i, respectively. This is expressed by

$$k_i^h \sim Binomial(h_i, t^s) \tag{14}$$

$$k_i^f \sim Binomial(f_i, t^n), \tag{15}$$

where t^s and t^n are the number of signal and noise trials presented in the experiment. As in Buntine (1994) this graphical model uses shading to differentiate observed from unobserved variables, and double border to represent deterministic variables. A new aspect of this representation is that shape is used to differentiate continuous from discrete variables. Like Lunn et al. (2000), Lee (2008) indicated the unit index and the total number of repetitions, but he did not follow the former in using double-edged arrows to denote a deterministic dependency between variables. The distributions are not represented in the graphical model.

After Lee's (2008) article the plate notation gained popularity within mathematical psychology, and it has been subsequently used in a number of studies (e.g., Ahn, Krawitz, Kim, Busemeyer & Brown 2011, Bridwell, Hecker, Serences & Srinivasan 2013, Dyjas, Grasman, Wetzels, van der Maas & Wagenmakers 2012, Lee & Newell 2011, Lodewyckx, Kim, Lee, Tuerlinckx, Kuppens & Wagenmakers 2011, Orhan & Jacobs 2013, Scheibehenne, Rieskamp & Wagenmakers 2013, Steyvers, Lee & Wagenmakers 2009, van Ravenzwaaij, Moore, Lee & Newell 2014), and in the book "Bayesian Cognitive Modeling: A Practical Course" by Lee & Wagenmakers (2014).

5.2. Kruschke's (2010) Hierarchical Bayesian Graphical Models with Distribution Pictograms

Kruschke (2010*a*, 2010*b*) introduced a new graphical representation of hierarchical Bayesian models in his book "Doing Bayesian Data Analysis: A tutorial with R and BUGS". He explained the purpose of introducing his graphical representation in a post of October 2013 in his blog (doingbayesiandataanalysis.blogspot.com). He stated that the directed acyclic graphs are incomplete and/or confusing for him, and that designing graphical models with the pictograms of the distributions help him explaining, inventing and programming models. As shown in Figure 5 this graphical model does not use plates, the variables are not represented by nodes but by letters, and pictograms with a prototypical distribution shapes are used. Moreover, the arrows are accompanied by a $\sim$ sign indicating a stochastic dependency, and a $=$ sign denoting deterministic dependency. The repetition is represented by three dots and the index (e.g., $\ldots i$), and when variables repeat over two different sets of units it is indicated with the conditional notation (e.g., $j \mid i$).

Moreover, unlike in the graphs with plate notation, the equations to calculate the deterministic variables are included in the model. The graphical model in Figure 5 was presented by Kruschke in his blog (which differs from the models in his book in which only "$\ldots$" without the index is used to represent repetition) with the purpose of comparing his novel graphical model with the more established graphical models with plate notation. It represents a model described by Spiegelhalter, Thomas, Best & Gilks (1996), in which $y_{j|i}$ denotes the weight of rat i at day after birth j. This variable is generated by the normal distribution so that:

$$y_{j|i} \sim Normal(w_{j|i}, \lambda), \tag{16}$$

where $w_{j|i}$ is the mean and λ is the precision. $w_{j|i}$ is deterministically calculated by

$$w_{j|i} = \phi_i + \xi_i x_{j|i}, \tag{17}$$

where ϕ_i is the intercept and ξ_i is the slope, and both come from normal distributions with means κ and ζ,; and precision δ and γ, respectively so that:

$$\phi_i \sim Normal(\kappa, \delta) \tag{18}$$

$$\xi_i \sim Normal(\zeta, \gamma) \tag{19}$$

The mean parameters come from a normal distribution with priors M and H, and the precisions come from a gamma distribution with parameters K and I.

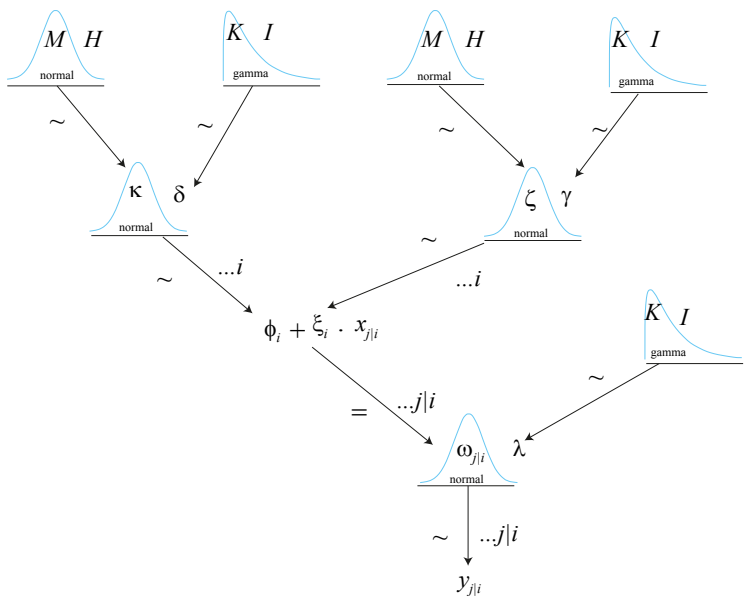

FIGURE 5: Adaptation of the type of graphical model representation introduced by Kruschke (2010*a*, 2010*b*). Stochastic variables are represented by small letters which are generated by a distribution (represented by pictograms) with the corresponding parameters. Stochastic variables are represented with small letters, and their values are computed by the corresponding equation. Arrows with the sign $\sim$ denote stochastic dependencies, and arrows with the sign= represent deterministic dependencies. The pictograms were generated using Bååth's (2013) template for LibreOffice.

This graphical representation was subsequently used by Kruschke and his colleagues (e.g., Kruschke 2013, Kruschke, Aguinis & Joo 2012). Bååth (2013) created a script in R (R Development Core Team 2013) to produce Kruschke's distribution pictograms and a template for the LibreOffice (Document Foundation, 2013)

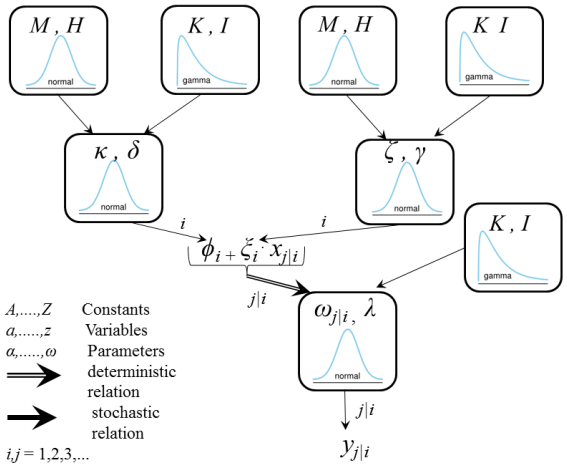

FIGURE 6: Adaptation of Schneider's (2013) representation of Kruschke's (2010) style graphical models. We generated the pictograms with Bååth's (2013) template for LibreOffice. The shapes surrounding the distributions are not plates, they only emphasise to which distribution the parameters belong to.

free software to draw the graphical diagrams with the distribution pictograms. Schneider (2013) presented a graphical model using Bååth's pictograms. This representation differs from that of Kruschke in that boxes were added around the distributions and double-edge arrows were used to denote deterministic relations. (Note that instead of the single-edged arrow used in Figure 6, Schneider used a curly arrow), and the deterministic equation is surrounded by a brace. Summing up, Kruschke's (2010*a*, 2010*b*) graphical model puts emphasis on the type of distribution. The repetition is not represented by plates, rather three dots and the indices of the variables are presented to denote the repetition. Another important aspect of this graphical model is that the parameters that belong to the same distribution are presented contiguously in space. (Note that this is even more emphasized in Schneider's representation, in which the parameters that form a distribution are surrounded by a box). As indicated by Kruschke the inclusion of distribution pictograms plays a heuristic role for the invention, explanation and programming of models. A possible problem with this approach is that sometimes a deterministic variable comes from a long relation between other variables, which could use a lot of space and make the graph difficult to understand.

6. Discussion

Both the graphical models with plates and the graphical models with distribution pictograms share the importance of representing the dependencies between variables by using arrows. The former emphasise the independent repetition of the variables, and the latter emphasises the probabilistic distributions. As mentioned earlier, Buntine's (1994) graphical models contained both plates and the names

(instead of the pictograms) of the distributions. However, Buntine only presented graphical models with one or two plates, and the larger the number of plates the smaller the space to include information about the distribution.

Moreover, complex models are also difficult to represent with graphical models with plates. For example, Figure 6.3 shows a hierarchical Bayesian model presented by Lee (2008). The complexity of the model led the author to use three i plates and two j plates. Although this solution is flawless from the formal point of view, it reduces the heuristic value of the graphical models with plates. Given that there is no differentiation between plates (i.e., they all have the same format) the only graphical difference between, say the plate i and the plate j is that they are different plates. Therefore, adding more plates for denoting repetition over the same set of units might be confusing for some researchers. One obvious solution to this problem is to colour code or shape code the plates. That is, to use a different colour (or shape) for each type of plate. Based on this summary we aimed at developing a type of graphical model that incorporates both the distribution pictograms and the plate notation, that solves the problems identified in the discussion, and that it has heuristic value to help researchers invent, explain and program hierarchical Bayesian models.

6.1. Hierarchical Bayesian Graphical Models with Distribution Pictograms and Mini-Plates

The first ingredient of our proposal is the replacements of plates by colour-coded mini-plates. (Note that in the paper version of this article we use shadings; please refer to the link provided below to see a colour version of the graphical models). As explained above, the purpose of colour-coding is to graphically differentiate between different types of plates. While we were developing the idea of using colour to differentiate between plates we realise that if the colour indicates the set of units over which the repetition occurs then the plates might not be necessary. Thus, we originally developed graphical models without plates in which we colour coded the nodes. However, because colour (or shading) is already used to code for variable type (i.e., observed vs. unobserved) this implementation was unsatisfactory.

That led us to develop the idea of colour coded mini-plates, and use them to surround each node. However, this was also unsatisfactory because the mini-plates with the nodes occupied too much space. In parallel we were also considering the use of 3D representations in order to make the graphical models more attractive. When we were trying different ways of using 3D nodes we found out that in some cases the combination of 3D shapes occupy less space than 2D shapes because the former allow more flexibility to locate the nodes in space. Thus, we realised that using 3D nodes was also a good idea to save space, and then we came up with the idea of using 3D rotated mini-plates under the nodes, instead of surrounding the nodes.

Unfortunately, in order to trigger the perception of 3D nodes it is necessary to add some shading to the nodes, which, as indicated by a reviewer, was very

distracting. Therefore, we used 3D mini-plates and 2D nodes. Note that in the link provided below we used 3D nodes with colour because the shading in some colours is not distracting. Figure 7 shows the mini-plate version of the bottom right part of the graphical model presented in Figure 4. (Note that this is a black-grey-white version of the graphical model. A colour version of Figure 7, as well as Figure 8 and Figure 11, could be accessed in this link: `http://dx.doi.org/10.6084/m9.figshare.1020148`. Moreover, a Microsoft PowerPoint template to create the figures could be accessed in this link: `http://dx.doi.org/10.6084/m9.figshare.1020020`) As in 4, the nodes of the variable k_{fi} and t_n are shaded because they are observed variables, and they have the shape of a square to indicate that they are discrete variables. Likewise, the f_i node is unshaded to denote that it is an unobserved variable, and it is a circle to indicate that it is a continuous variable. Like in Figures 3 and 6, we adopted the double-arrow to indicate a deterministic dependency. Thus, the f_i node does not have a double border as in Figure 4. The grey mini-plates under the f_i and k_{fi} indicate that they repeat over $i = 1$ to N units. Mini-plates play exactly the same role as the plates; that is, they indicate that the variables within them are repeated over the number the units indicated in the plates or mini-plates.

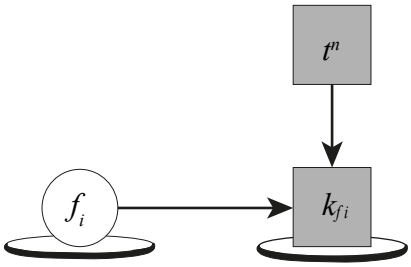

FIGURE 7: Graphical model representation with mini-plates representing the bottom right hand side part of the model illustrated in Figure 4. The white plates under f_i and k_{fi} indicate repetition over units $i = 1$ to N. The dark grey square denotes an observed discrete variable and the unshaded circle indicates unobserved continuous variable. The stochastic dependencies are indicated by arrows.

Having mini-plates provides flexibility to the modeler to where to position the nodes in space, and, as a consequence, it makes easier to incorporate Kruschke's distribution pictograms. Figure 8 shows the mini-plate version of the bottom part of the graphical model presented in Figure 5. As in Figure 5, the mini-plate version incorporates the pictogram of the normal distribution with the parameters represented with unshaded circles. Note that two plates are located under y_{ij}: The grey i plate and the diagonally stripped j plate. Likewise the normal distribution and the ω_{ij} are on top of an i plate and a j plate, indicating that the process that generates y_{ij} repeats over the i and j units. However the λ node is not on top of any plate because the generating process uses only one value of λ.

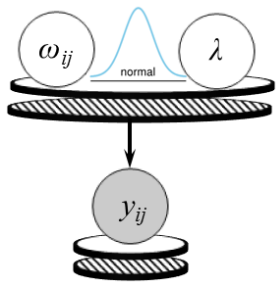

FIGURE 8: Graphical model with mini-plates and distribution pictograms representing the bottom part of Figure 5. The white plate under y_{ij} and ω_{ij} indicate that these variables repeat over $i = 1$ to N units. The diagonally stripped plates under the same variables denote their repetition over $j = 1$ to M units. λ is a parameter of the normal distribution, and given that it is not located on top of plates, it does not repeat itself.

6.2. Comparison Between Three Graphical Model Representations of a Hierarchical Bayesian Model

Having presented the mini-plate with distribution pictograms, we now discuss whether this type of graphical representation is capable of representing complex models, and how it compares with the other two types of graphical models. For this purpose we present here a complex hierarchical Bayesian model with plate notation developed by Lee (2008, pages 5 to 8), based on Nosofsky (1986) generalised context model (GCM) of category learning.

This model aims at explaining how people learn to categorise unknown stimuli into two categories in experiments in which the researcher uses different category structures (see more details in Lee, 2008). In those experiments the researcher assigns one fourth of the total number of stimuli to one category (i.e., category A), one fourth to the other category (i.e., category B), and one half is not assigned to any category. The model utilises the multidimensional scaling (MDS) representation of stimulus similarity developed by Shepard (1962). Stimuli are represented as points (p) in a D-dimensional space. In Figure 9 p_{ix} denotes a coordinate value of stimulus i in dimension x. The surrounding plates indicates that, in this example, there are N stimuli and $x = 2$ dimensions. Lee (2008) assigned the following prior probability to p_{ix} (not shown in the graph):

$$p_{ix} \sim Uniform(-\delta, \delta); \delta > 0, \tag{20}$$

d^2_{ij} denotes the squared psychological distance between all the possible pairs of N stimuli, where i denotes the first stimulus and j the second stimulus of the pair. The node is surrounded by the i and j plates to denote the repetition over all the possible pairs of stimuli. The squared psychological distance between stimuli is determined by:

$$d^2_{ij} = w(p_{i1} - p_{j1})^2 + (1 - w)(p_{i2} - p_{j2})^2, \tag{21}$$

where w is the relative attention paid to the first stimulus over the second stimulus, and has the following prior distribution (not shown in the graph):

$$w \sim Uniform(0, 1). \tag{22}$$

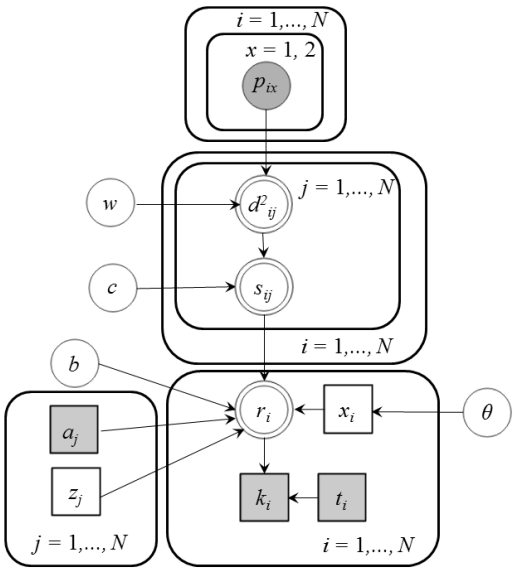

FIGURE 9: Adaptation of the graphical model with plates presented by Lee (2008, page 6, Figure 5). The notation is the same as in Figure 4.

The similarity between each pair of stimuli is given by:

$$s_{ij} = \exp[-(cd_{ij})^2], \tag{23}$$

where c is a generalisation gradient parameter, with the following prior (not shown in the graph):

$$c^2 = Gamma(\varepsilon, \varepsilon), \tag{24}$$

where $\varepsilon = .001$.

The number of times the ith stimulus is chosen as a member of category A out of t_i trials is denoted by k_i, which follows a binomial distribution:

$$k_i \sim Binomial(r_i, t_i), \tag{25}$$

where r_i is the probability of stimulus i being chosen as a member of category A. This probability is determined by the similarities between stimuli (s_{ij}), a response bias b, how the stimuli were assigned to the categories by the researcher (i.e., indicator variables a_j, z_j), and a third indicator variable x_i. The prior distribution of the response bias b (not shown in the graph) is:

$$b \sim Uniform(0, 1). \tag{26}$$

a_j denotes the known assignment of the jth presented stimulus, which ranges over N/2 such stimuli, and z_j indicates the latent assignment of the jth unassigned stimulus, ranging over the N/2 such stimuli (see Lee, 2008, for a more detailed explanation of stimuli assignment structures), with the following prior:

$$z_j \sim Bernoulli(1/2). \tag{27}$$

The x_i is incorporated in the graph in order to compare a model that uses the latent stimulus assignment z_j with a simpler model without this latent variable. x_i indicates for each stimulus i whether the response probability r_i uses a_j and z_j (i.e., $x_i = 1$) or only a_j (i.e., $x_i = 0$). It is assumed that all the indicators x_i support either model following a fixed underlying rate of use (i.e., θ), which is given by the following prior:

$$\theta \sim Uniform(0, 1). \tag{28}$$

The posterior rate of use provides information on how well each model accounts for the categorisations of the participants (for a more detailed explanation see Lee, 2008, page 8). Finally, the probability for the ith stimulus to be classified as a member of category A is given by:

$$r_i = \frac{b \sum_{a \epsilon A} s_{ia}}{b \sum_{a \epsilon A} s_{ia} + (1-b) \sum_{a \epsilon B} s_{ia}}, \tag{29}$$

if x_i is 0.

$$r_i = \frac{b[\sum_{a \epsilon A} s_{ia} + \sum_{z \epsilon A} s_{iz}]}{b[\sum_{a \epsilon A} s_{ia} + \sum_{z \epsilon A} s_{iz}] + (1-b)[\sum_{a \epsilon B} s_{ia} + \sum_{z \epsilon B} s_{iz}]}, \tag{30}$$

if x_i is 1.

Figure 9 shows Lee's (2008) graphical model with plates, 10 presents our best attempt to represent the same model with the graphical model with distribution pictograms proposed by Kruschke (2010*a*), and Figure 11 depicts the graphical model with mini-plates and distribution pictograms, as proposed in this article. Regarding the amount of information presented, Kruschke's graphical model presents more information (equations, distributions, dependencies and repetitions) than the mini-plate graphical model (distributions, dependencies and repetitions) and the plate graphical model (dependencies and distributions). However, it seems to us that the inclusion of equations in Kruschke's graphical model attempts against the heuristic value of the model by cluttering the space with too much information. Possibly, Kruschke's style graphical model should include equations only when the models are simple and exclude them when the models contain more than one or two equations and/or when the equations are large.

The mini-plate graphical representation seems to strike a balance between amount of information and use of space. By using the concept of plates in a flexible way (i.e., mini-plates), combining them with Kruschke's distribution pictograms, and using 3D shapes it has a strong visual appeal. However, this comparison is not completely fair for the other graphical representations. As we mentioned

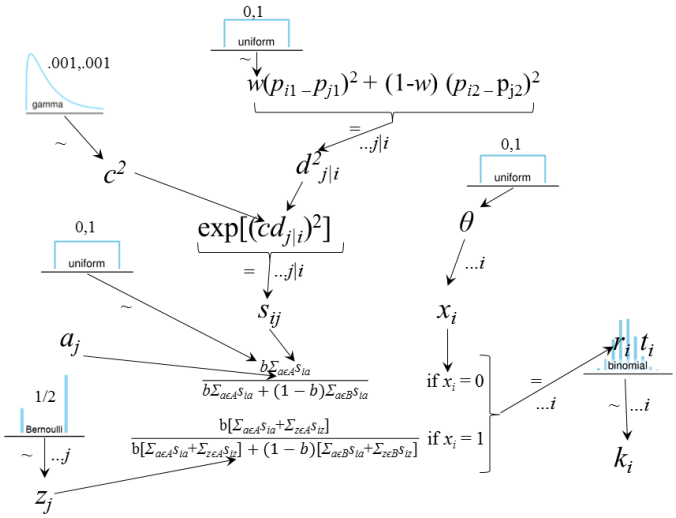

FIGURE 10: Graphical model with Kruschke's style distribution pictograms of the hierarchical Bayesian model presented by Lee (2008, pages 5 to 8). The notation is the same as in Figure 5.

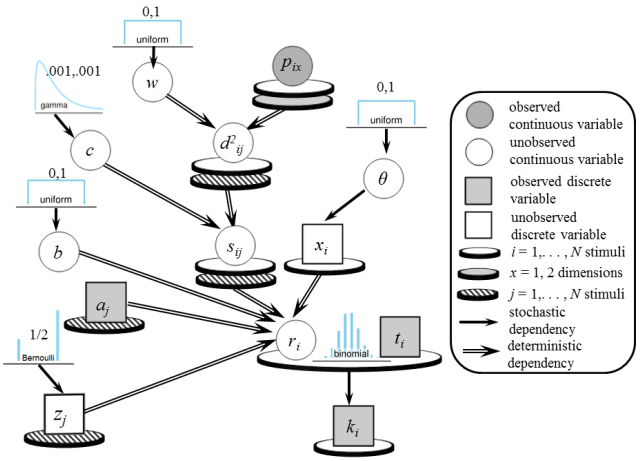

FIGURE 11: Graphical model with mini-plates of the hierarchical Bayesian model presented by Lee (2008, pages 5 to 8). The notation is the same as in Figure 8.

above, Kruschke's graphical models could be improved by not including equations in complex models, and we might not have done the best to present the best possible representation. Moreover, there are a variety of plate graphical models that we are not displaying in this article. For example, van Ravenzwaaij et al. (2014)

presented a plate graphical model following Lee's (2008) notation, but in the right hand side of the figure they added both the distributions and the equations in text format. Allegedly, this might be a better representation than the mini-plate graphical models because it combines graphical aspects with text, rather than being too graphical.

7. Conclusion

We discussed two types of hierarchical Bayesian graphical models used in psychology –Lee's (2008) plate graphical models and Kruschke's (2010*b*) graphical models with distribution pictograms. We proposed a third type– graphical models with colour-coded mini-plates– as an attempt to combine the positive aspects of the other two types of graphical models, and we presented a preliminary intuitive analysis of the informative value of this new graphical respresentation. We believe our proposal provides an important contribution to the field, but our intuitive evaluation should be confirmed by empirical evidence. Statistical cognition is a flourishing area of research in psychology (e.g., Beyth Marom, Fidler & Cumming 2008), that investigates how people understand statistical information. Our proposal could be followed up by a research that aims at elucidating which type of graphical model leads to a better understanding of hierarchical Bayesian models.

[Recibido: mayo de 2014 — Aceptado: septiembre de 2014]

References

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R. & Brown, J. W. (2011), 'A model-based FMRI analysis with hierarchical Bayesian parameter estimation', *Journal of Neuroscience, Psychology, and Economics* **4**(2), 95–110.

Bauer, M. I. & Johnson-Laird, P. N. (1993), 'How diagrams can improve reasoning', *Psychological Science* **4**(6), 372–378.

Beyth Marom, R., Fidler, F. & Cumming, G. (2008), 'Statistical cognition: Towards evidence-based practice in statistics and statistics education', *Statistics Education Research Journal* **7**(2), 20–39.

Bolstad, W. M. (2007), *Introduction to Bayesian Statistics*, 2 edn, Wiley Jobs in Hoboken, New York.

Bridwell, D. A., Hecker, E. A., Serences, J. T. & Srinivasan, R. (2013), 'Individual differences in attention strategies during detection, fine discrimination, and coarse discrimination', *Journal of Neurophysiology* **110**, 784–794.

Bååth, R. (2013), *Distribution diagram R scripts.* *https://github.com/rasmusab/distribution_diagrams

Buntine, W. L. (1994), 'Operations for learning with graphical models', *Journal of Artificial Intelligence Research* **2**, 159–225.

Dienes, Z. (2011), 'Bayesian versus orthodox statistics: Which side are you on?', *Perspectives on Psychological Science* **6**, 274–290.

Dyjas, O., Grasman, R. P., Wetzels, R., van der Maas, H. L. & Wagenmakers, E. (2012), 'A Bayesian hierarchical analysis of the name-letter effect', *Frontiers in Psychology* **3**(334), 1–14. doi: 10.3389/fpsyg.2012.00334.

Edwards, W., Lindman, H. & Savage, L. J. (1963), 'Bayesian statistical inference for psychological research', *Psychological Review* **70**, 193–242.

Gilks, W. R., Thomas, A. & Spiegelhalter, D. J. (1994), 'A language and program for complex Bayesian modelling', *The Statistician* **43**, 169–177.

Good, I. J. (1980), 'Some history of the hierarchical Bayesian methodology', *Trabajos de Estadística y de Investigación Operativa* **31**(1), 489–519.

Gray, K. & Wegner, D. M. (2013), 'Six guidelines for interesting research', *Perspectives on Psychological Science* **8**(5), 549–553.

Griffiths, T. L., Kemp, C. & Tenenbaum, J. B. (2008), Bayesian models of cognition, *in* R. Sun, ed., 'The Cambridge Handbook of Computational Psychology', Cambridge University Press, Cambridge, New York, pp. 59–100.

Heit, E. & Rotello, C. (2005), Are there two kinds of reasoning?, *in* B. G. Bara, L. W. Barsalou & M. Bucciarelli, eds, 'Proceedings of the 27th Annual Conference of the Cognitive Science Society', Lawrence, Erlbaum, Mahwah, New Jersey.

Jordan, M. I. (2004), 'Graphical models', *Statistical Science* **19**, 140–155.

Koller, D. & Friedman, N. (2009), *Probabilistic Graphical Models: Principles and Techniques*, MIT press, Cambridge, MA.

Koller, D., Friedman, N., Getoor, L. & Taskar, B. (2007), Graphical models in a nutshell, *in* L. Getoor & B. Taskar, eds, 'Introduction to statistical relational learning', MIT Press, Cambridge, Massachusetts.

Kruschke, J. K. (2010*a*), *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*, Academic Press, Burlington, MA.

Kruschke, J. K. (2010*b*), 'What to believe: Bayesian methods for data analysis', *Trends in Cognitive Science* **14**, 293–300.

Kruschke, J. K. (2013), 'Bayesian estimation supersedes the t test', *Journal of Experimental Psychology: General* **142**, 573–603.

Kruschke, J. K., Aguinis, H. & Joo, H. (2012), 'The time has come: Bayesian methods for data analysis in the organizational sciences', *Organizational Research Methods* **15**, 722–752.

Lee, M. D. (2008), 'Three case studies in the Bayesian analysis of cognitive models', *Psychonomic Bulletin and Review* **15**, 1–15.

Lee, M. D. (2011), 'How cognitive modeling can benefit from hierarchical Bayesian models', *Journal of Mathematical Psychology* **55**, 1–7.

Lee, M. D. & Newell, B. R. (2011), 'Using hierarchical Bayesian methods to examine the tools of decision-making', *Judgment and Decision Making* **6**, 832–842.

Lee, M. D. & Wagenmakers, E.-J. (2005), 'Postscript: Bayesian statistical inference in psychology: Comment on Trafimow (2003)', *Psychological Review* **112**(3), 662–668.

Lee, M. D. & Wagenmakers, E.-J. (2014), 'Bayesian cognitive modeling: A practical course', *New York: Cambridge University Press* .

Lodewyckx, T., Kim, W., Lee, M. D., Tuerlinckx, F., Kuppens, P. & Wagenmakers, E. J. (2011), 'A tutorial on Bayes factor estimation with the product space method', *Journal of Mathematical Psychology* **55**, 331–347.

Lunn, D. J., Thomas, A., Best, N. & Spiegelhalter, D. (2000), 'A Bayesian modelling framework: Concepts, structure, and extensibility', *Statistics and Computing* **10**, 325–337.

Nosofsky, R. M. (1984), 'Choice, similarity, and the context theory of classification', *Journal of Experimental Psychology: Learning, Memory and Cognition* **10**, 104–114.

Nosofsky, R. M. (1986), 'Attention, similarity, and the identification–categorization relationship', *Journal of Experimental Psychology: General* **115**, 39–57.

Orhan, A. E. & Jacobs, R. A. (2013), 'A probabilistic clustering theory of the organization of visual short-term memory', *Psychological Review* **120**, 297–328.

Pearl, J. (2009), 'Causal inference in statistics: An overview', *Statistics Surveys* **3**, 96–146.

R Development Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Scheibehenne, B., Rieskamp, J. & Wagenmakers, E.-J. (2013), 'Testing adaptive toolbox models: A Bayesian hierarchical approach', *Psychological Review* **120**, 39–64.

Schneider, T. (2013), *Diagram for Hierarchical Models.*
*https://github.com/tinu-schneider/DBDA_hierach_diagram

Shepard, R. N. (1962), 'The analysis of proximities: Multidimensional scaling with an unknown distance function. I', *Psychometrika* **27**(2), 125–140.

Shepard, R. N. (1980), 'Multidimensional scaling, tree-fitting, and clustering', *Science* **210**, 390–398.

Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996), *BUGS 0.5* Examples Volume 2 (version ii)*, 2 edn, MRC Biostatistics Unit.

Stenning, K. & Oberlander, J. (1995), 'A cognitive theory of graphical and linguistic reasoning: Logic and implementation', *Cognitive Science* **19**(1), 97–140.

Steyvers, M., Lee, M. D. & Wagenmakers, E.-J. (2009), 'Bayesian analysis of human decision-making on bandit problems', *Journal of Mathematical Psychology* **53**(3), 168–179.

van Ravenzwaaij, D., Moore, C. P., Lee, M. D. & Newell, B. R. (2014), 'A hierarchical Bayesian modeling approach to searching and stopping in multi-attribute judgment', *Cognitive Science* **38**, 1384–1405.

Wagenmakers, E. (2007), 'A practical solution to the pervasive problems of p values', *Psychonomic Bulletin and Review* **14**, 779–804.

Wickens, C. D., Merwin, D. H. & Lin, E. L. (1994), 'Implications of graphics enhancements for the visualization of scientific data: Dimensional integrality, stereopsis, motion, and mesh', *Human Factors* **36**(1), 44–61.