Bayesian Analysis of Rank Data With Application to Primate Intelligence Experiments

Valen E. JOHNSON, Robert O. DEANER, and Carel P. VAN SCHAIK

A model for analyzing rank data obtained from multiple evaluators, possibly using different ranking criteria, is proposed. The model is specified hierarchically within the Bayesian paradigm and includes parameters that represent the probabilities that two items are assigned equal rankings. Also included are parameters that account for the relative precision of rankings obtained from distinct evaluation schemes. The model is illustrated through a meta-analysis of rank data collected to compare the cognitive abilities of various primate genera.

KEY WORDS: Hierarchical Bayes; Markov chain Monte Carlo; Meta-analysis; Paired-comparisons; Thurstonian model.

1. INTRODUCTION

In this article we propose a Bayesian latent variable model for analyzing rank data obtained from several observers or studies. The resulting model is closely related to Thurstonian models for rank data (Thurstone 1927, 1931) as developed by, among others, Daniels (1952), Stern (1990), and Böckenholt (1992, 1993) and summarized, along with other methods for analyzing rank data, by Marden (1995). The model is also related to classical paired-comparison models (e.g., Bradley and Terry 1952; Bradley 1984; David 1988) and the more recent work on paired-comparison models for time-varying data by Glickman and Stern (1998) and Glickman (1999). From a Bayesian perspective, the model is closely related to multirater ordinal data models as described by Johnson (1996, 1997) and Johnson and Albert (1999).

The major innovations of this model are the inclusion of parameters to accommodate ties and a hierarchical structure that facilitates the estimation of variances and correlations arising from related studies. (See Davidson 1970 for an alternative formulation for tied data in standard paired-comparison models.) In addition, posterior probabilities concerning the relative values of ranked items and other model parameters are easily assessed using Markov chain Monte Carlo (MCMC) methods.

To illustrate the features of the methodology, the model is defined in the context of a meta-analysis of historical data collected to rank nonhuman primate taxa according to their "intelligence." (We use the term "intelligence" as a theoretical construct meaning that some cognitive abilities are relatively domain general or applicable to many situations. The use of the term does not imply that all processing can necessarily be reduced to a single unidimensional axis. For further discussion of this issue, see Deaner, van Schaik, and Johnson 2001.) Although the extent and nature of taxonomic differences in cognitive abilities is of long-standing interest for evolutionary biologists, few statistical tools are available for assessing performance differences across multiple problem types (Balda, Kamil, and Bednekoff 1996; Lefebvre and Giraldeau 1996; Rumbaugh, Savage-Rumbaugh, and Washburn 1996; Tomasello and Call 1997).

Consequently, the proposed model should ultimately permit substantial progress toward elucidating the organization and evolution of animal cognition, especially regarding the question of whether and to what extent domain-general abilities exist in nonhuman primates.

The meta-analysis considered here includes several dozen studies, which are grouped into *paradigms* and *procedures*. Paradigms are general types of "intelligence" tests, whereas procedures refer to different methodologies used in investigating these paradigms. Taxonomic rankings within a procedure usually derive from a single study, but in some cases, two or more studies with extremely similar methods were conducted and hence were combined. To familiarize readers with these studies, we provide a brief description of each paradigm. Details on the procedures within the paradigms are provided in earlier work (Deaner et al. 2001). Rank data for this meta-analysis appear in Table 1. The following paradigms were included in the meta-analysis:

- **Detour**. In *detour problems*, the subject is required to manually move an object through a spatial field containing obstacles. Detour problems investigate the ability to form and act on spatial representations.
- **String**. In *patterned string* problems, the subject is shown an array of interlaced strings, one of which is tethered to a desirable food. The subject is allowed to pull only one of the strings, and hence must determine which string is actually attached to the reward. Patterned string problems also investigate the ability to form and act on spatial representations.
- **Displace**. In studies of *invisible displacement*, the subject views an object being placed into a container that is subsequently moved behind one or more barriers. The subject is then shown the empty container. If the subject searches only the barriers behind which the container passed, this indicates that the subject can track the spatial movements of unperceived objects.
- **Tool**. *Tool use*, in which the subject must move an intermediate object (a tool) in relation to another object or substrate, involves aspects of causal reasoning, spatial representation, and motor coordination.
- **Discriminate**. In the *object discrimination learning set* paradigm, the subject is first confronted with the prob-

Valen E. Johnson is a Technical Staff Member, Statistical Sciences, MS F600, Los Alamos National Laboratory, Los Alamos, NM 87545 (E-mail: *valen@lanl.gov*), and is on leave from the Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708, where he is a Professor. Robert O. Deaner is a Research Associate, Department of Neurobiology, Duke University Medical Center, Durham, NC 27710. Carel P. van Schaik is a Professor, Department of Biological Anthropology, Duke University, Durham, NC 27708.

Table 1. Studies Included in the Meta-Analysis of Primate Intelligence

Paradigm	Detour			String	Displace			Tool				Discriminate							Reversal					Odd			Sort		Delay	
Procedure / Genus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Pan	1	1		2				1.5	1.5						2								1.5	1			1	1	5	
Pongo				1		1	1		1.5						4								3						1	1
Gorilla				4	1			1.5							3								1.5						8.5	2.5
Hylobates															5						2		6.5						5	
Presbytis														1																
Macaca	2	2	1	6.5	2.5		2	3		1		1.5	1	2	1	1		1	2.5	1			4	2	3		2.5	2	5	2.5
Mandrillus				6.5																									5	
Cercopithecus			2	6.5																		1	5		6				5	
Papio				10																									2	
Cercocebus				6.5																									8.5	
Miopithecus																2							11							
Ateles				3							1								1											
Aotus																					3									
Lagothrix			3																						1				11	
Callithrix										3								2												4
Saimiri										2	3										4				4	2				
Cebus			4	9	2.5	2		1.5	3	2	2	1.5	2						2.5	2	1		6.5		2	1	2.5		11	
Eulemur																						2.5								
Nycticebus																	2													
Phaner																							9							
Varecia																													11	
Microcebus																	3						10							
Galago												3					1						8		5					
Lemur			5									3								3		2.5								

NOTE: Entries denote rankings obtained under each procedure (columns) for genera considered in the procedure. The first grouping within the table consist of great apes, the second New World monkeys, the third group Old World monkeys, the fourth New World monkeys, and the last Prosimians. Hylobates are not so categorized. Original data on which the rankings are based have been given in the following references: 1, Davis, McDowell, and Nissen (1957); 2, McDowell and Nissen (1959); 3, Davis and Leary (1968); 4, Harlow and Settlage (1934), Finch (1941), Reisen, Greenberg, Granston, and Fantz (1953), Fischer and Kitchener (1965), Balasch, Sabater-Pi, and Padrosa (1974); 5, Natale, Antinucci, Spinozzi, and Poti (1986); 6, de Blois, Novak, and Bond (1998); 7, de Blois, Novak, and Bond (1999); 8, Natale and Antinucci (1989b); 9, Visalberghi, Rumbaugh, and Fragaszy (1995); 10, Miles and Meyer (1956); 11, Shell and Riopelle (1958); 12, Stevens (1965); 13, Stevens (1965); 14, Manocha (1967); 15, Rumbaugh and McCormack (1967); 16, Schrier (1972); 17, Ohta (1983), Ohta, Ishida, and Matano (1984, 1987); 18, Meyer (1951); 19, Crawford (1962); 20, Stevens (1965); 21, Gossette and Inman (1966); Gossette and Slonim (1969), Gossette (1970); 22, Rumbaugh and Arnold (1971); 23, Rumbaugh (1997); 24, Strong and Hedges (1966); 25, Davis, Leary, Stevens, and Thompson (1967); 26, Thomas and Boyd (1973); 27, Garcha and Ettlinger (1979); 28, Tinklepaugh (1932); 29, Harlow, Uehling, and Maslow (1932), Maslow and Harlow (1932); 30, Miles (1957a,b).

lem of discriminating between two objects. One of the objects is arbitrarily designated correct, and the subject is rewarded for selecting it. The subject is given several trials under these conditions and learns to consistently make the correct choice. The learning set phenomenon refers to the observation that if the subject is given another discrimination problem, with two novel stimuli, then it will tend to learn this second problem more quickly than it did the first one. The degree to which the subject "learns how to learn" is thought to reflect on its ability to form and use abstract rules or "hypotheses."

- **Reversal**. The *reversal learning* paradigm investigates the ability to reverse a previously learned discrimination. Most commonly, over the course of several trials, the subject learns to make one object discrimination to get a reward (e.g., picking one object rather than another). Then, without warning, the values of the objects change so that the previously unrewarded object is rewarded for a run of trials.

- **Odd**. The *oddity* paradigm addresses the ability to use a relational or abstract concept. In most studies, the subject is simultaneously provided with three visual stimuli, two of which are identical and one that differs; the subject is rewarded for choosing the differing or odd stimulus.

- **Sort**. The *sorting* paradigm examines the ability to form abstract concepts and to use them to categorize stimuli accordingly. Tests usually require the subject to place similar objects in the same container.

- **Delay**. The *delayed response* paradigm investigates the subject's memory. In most studies, the subject observes a reward being hidden in one of two spatial locations, there is a delay, and then the subject is allowed to search one of the locations. The questions of interest are "For any given time interval, what percentage of first searches are correct?" and "What is the maximum delay at which a subject can still score above chance?"

2. MODEL SPECIFICATION

A general question that arises in the interpretation of rank data obtained from multiple paradigms (i.e., studies or observers) is whether or not there exists a single underlying trait that can explain each paradigm's rankings. This question is especially important in the meta-analysis considered here because of the controversy surrounding the existence of domain-general abilities (intelligence) and the fact that several distinct paradigms were used. Thus it is important that the statistical model used for the analyses of these data include parameters to represent paradigm–genus biases (or, in more general terms, biases of rankings obtained from studies conducted under paradigms for which particular items are differentially assigned higher or lower rankings). These parameters aid the detection of biases associated with given paradigms when they exist, and also permit such effects to be disregarded when they are not supported by data. When present, the inclusion of paradigm–genus bias parameters provides a simple mechanism for accounting for the correlation between rankings obtained from procedures within the same paradigm.

With these considerations in mind, we assume that the ranks obtained under each procedure are based on two under-

lying, continuous-valued latent variables: a global trait variable and a paradigm–genus bias effect. In the context of the primate intelligence study, the global trait variables represent each genus's underlying cognitive ability—here assumed to be unidimensional—and the paradigm-genus bias represents variations in the measurement of the global trait according to paradigm. Large values of paradigm-bias effects may be used to diagnose violations of the assumption of a unidimensional global trait variable. The combination of these variables is assumed to be observed with error.

To make these assumptions more precise, let $y_{i,j}$ denote the rank of the ith genus obtained from the jth procedure, and let $g(j)$ denote the paradigm to which the jth procedure belongs. We assume that $y_{i,j} > y_{k,j}$ implies $z_{i,j} > z_{k,j}$, where the latent procedure variable, $z_{i,j}$, can be expressed as

$$z_{i,j} = \theta_i + \eta_{i,g(j)} + \epsilon_{i,j}. \qquad (1)$$

In the context of primate intelligences, θ_i denotes the global intelligence measure of the ith genus, $\eta_{i,g(j)}$ denotes the paradigm–genus bias effect, and $\epsilon_{i,j}$ is a random error. Let K denote the number of paradigms under which procedures are defined, and let I denote the total number of genera ranked. To establish a measurement scale for the latent variables, we assume that the variables θ_i are independent and identically distributed a priori as standard normal variables. The paradigm–genus bias effects are also assumed to be independently distributed according to a normal distribution, but with mean 0 and precision $\tau_{g(j)}$. The precision parameter $\tau_{g(j)}$ is assumed to be constant for procedures within the same paradigm. The random errors $\epsilon_{i,j}$ are assumed to be independently distributed according to normal distributions with mean 0 and precision γ_j, where γ_j is unique to procedure j. The primate intelligence study includes 9 paradigms, 30 procedures, and 24 genera.

In the second stage of the hierarchical model, we assume that the γ_j's are drawn independently from a gamma distribution with mean and variation parameters μ_1 and ν_1, and that the $\tau_{g(j)}$'s are drawn independently from a gamma distribution with mean μ_2 and variation parameter ν_2.

The final stage of the model comprises prior distributions on the hyperparameters μ_1, μ_2, ν_1, and ν_2. In this stage of the model, we reflect vague prior information concerning the values of the precision parameters in the second stage, while also maintaining propriety of the posterior and avoiding degenerate peaks in the posterior that can occur when a subset of precision parameters becomes arbitrarily large. This difficulty occurs when one or more procedure's ranks exactly agree with the estimated consensus ranking, a configuration that can easily arise in small studies.

A further consideration that arises in setting the prior distributions in the third stage of the model concerns the identifiability of both the paradigm–genus precision parameters, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$, and the procedure precision parameters, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)$. When paradigm–genus bias terms are uniformly small, or if only a small number of procedures are performed under each paradigm, differentiating paradigm–genus bias effects from procedure errors is problematic. This suggests a two-step fitting procedure in which support for including paradigm–genus biases in the model is first assessed by

fitting a model including such effects, and then, if necessary, refitting a reduced model without them.

An additional relevant consideration in specifying the third-stage model involves scaling the trait variables. Because the latent intelligence variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)$ are defined on a standard normal scale, procedure precisions smaller than 1 imply that the ranking procedures are relatively uninformative. Similarly, values of the paradigm–genus precision variables ($\boldsymbol{\tau}$) smaller than 1 imply substantial disagreement between paradigms. Because the prior means of the γ_j and τ_k are μ_1 and μ_2, and their prior variances are parameterized to be μ_1^2/ν_1 and μ_2^2/ν_2, we thus assume a common exponential prior distribution on μ_1 and μ_2 with parameters fixed so that the prior mean of each is 20 and the prior variance is 400. This prior has its mode at the origin but places substantial weight on values of μ_1 and μ_2 greater than 10.

The prior on the variation parameters ν_1 and ν_2 was chosen to be an inverse gamma distribution that placed 80% of its weight between .25 and 4. In other words, the coefficient of variation for the second-stage precision parameters was assigned an 80% probability of being between 1/2 and 2. The sensitivity of the model to these third-stage model assumptions is addressed in Section 4.

To model information contained in ties, let $z_{(i,j)}, i \in C_j$, denote the ordered values of the observed traits of genera ranked in procedure j, with C_j denoting the set of genera ranked in study j and C_j^* denoting the set C_j with the genus estimated as having the largest value of $z_{i,j}$ omitted. Let $y_{(i,j)}$ denote the corresponding ordered values of the observed ranks. Define

$$p_{(i,j)}(\kappa)$$
$$= \begin{cases} \exp[-(z_{(i+1,j)} - z_{(i,j)})/\kappa] & \text{if } y_{(i+1,j)} = y_{(i,j)} \\ 1 - \exp[-(z_{(i+1,j)} - z_{(i,j)})/\kappa] & \text{if } y_{(i+1,j)} \neq y_{(i,j)} \end{cases} \quad (2)$$

for a given value of the parameter κ. The function $p_{(i,j)}(\kappa)$ defines a parametric model for the probability that two genera are tied, given their latent procedure variables. The prior distribution on κ is assumed to be uniform on the positive real line.

Given the aforementioned assumptions, the sampling density assumed for the observed rank data is assumed to take the form

$$f(\mathbf{y} \mid \mathbf{z}, \kappa, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \alpha, \beta, \lambda)$$
$$= \prod_{j=1}^{J} \prod_{i \in C_j^*} p_{(i,j)}(\kappa) \, \mathrm{Ind}(y_{(i+1,j)} \geq y_{(i,j)}). \quad (3)$$

The normalizing constant of this sampling density is independent of κ. It then follows that the posterior distribution on the model parameters can be expressed as

$$p(\mathbf{z}, \kappa, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \alpha, \beta, \lambda \mid \mathbf{y})$$
$$\propto \prod_{j=1}^{J} \prod_{i \in C_j^*} p_{(i,j)}(\kappa) \, \mathrm{Ind}(y_{(i+1,j)} \geq y_{(i,j)})$$

$$\times \prod_{j=1}^{J} \prod_{i \in C_j} \mathrm{n}(z_{i,j}; \theta_i + \eta_{i,g(j)}, \gamma_j) \times \prod_i \mathrm{n}(\theta_i; 0, 1)$$

$$\times \prod_{j=1}^{J} \prod_{i \in C_j} \mathrm{n}(\eta_{i,g(j)}; 0, \tau_{g(j)}) \times \prod_{j=1}^{J} \mathrm{gamma}(\gamma_j; \nu_1, \nu_1/\mu_1)$$

$$\times \prod_{k=1}^{K} \mathrm{gamma}(\tau_k; \nu_2, \nu_2/\mu_2) \times \mathrm{gamma}(\mu_1; 1, \alpha)$$

$$\times \mathrm{gamma}(\mu_2; 1, \alpha) \times \mathrm{inv\text{-}gamma}(\nu_1; \delta, \beta)$$

$$\times \mathrm{inv\text{-}gamma}(\nu_2; \delta, \beta). \quad (4)$$

Here $(\alpha, \beta, \delta) = (.05, .65, 1.17)$, $\mathrm{n}(\cdot; a, b)$ denotes a normal density with mean a and precision b, $\mathrm{gamma}(\cdot; c, d)$ denotes a gamma density with shape and scale parameters c and d (mean c/d), and inv-gamma denotes the corresponding inverse-gamma density.

3. MARKOV CHAIN MONTE CARLO ESTIMATION OF MODEL PARAMETERS

The normal-gamma conjugate structure assumed for the first- and second-stage models makes implementation of a hybrid Gibbs–Metropolis sampling scheme straightforward. After the model is initialized so that the parameter configuration is consistent with the data (i.e., all precision parameters and κ are positive, and the estimated procedure values z_{ij} are consistent with the observed rank data), the following steps may be used to generate samples from the posterior distribution:

1. For $i = 1, \ldots, I$, sample θ_i from a normal distribution with mean c/d and precision d, where

$$c = \sum_{j:i \in C_j} \gamma_j(z_{ij} - \eta_{i,g(j)}) \quad \text{and} \quad d = 1 + \sum_{j:i \in C_j} \gamma_j.$$

2. For $k = 1, \ldots, K$ and $i = 1, \ldots, I$, sample $\eta_{i,k}$ from a normal distribution with mean c/d and precision d, where

$$c = \sum_{j:g(j)=k} \gamma_j(z_{ij} - \theta_i) \quad \text{and} \quad d = \tau_{g(j)} + \sum_{j:g(j)=k} \gamma_j.$$

3. For $j = 1, \ldots, J$, sample γ_j from a gamma distribution with shape c and scale d, where

$$c = \nu_1 + \sum_{i \in C_j} \frac{1}{2} \quad \text{and} \quad d = \frac{\nu_1}{\mu_1} + \sum_{i \in C_j} \frac{(z_{ij} - \theta_i - \eta_{i,g(j)})^2}{2}.$$

4. For $k = 1, \ldots, K$, sample τ_k from a gamma distribution with shape c and scale d, where

$$c = \nu_2 + \sum_{j:g(j)=k} \sum_{i \in C_j} \frac{1}{2} \quad \text{and} \quad d = \frac{\nu_2}{\mu_2} + \sum_{j:g(j)=k} \sum_{i \in C_j} \frac{\eta_{i,g(j)}^2}{2}.$$

5. Generate a candidate draw for μ_1, say μ_*, from an inverse gamma with shape parameter c and scale parameter d, where

$$c = 1 + J\nu_1 \quad \text{and} \quad d = \sum_{j=1}^{J} \nu_1 \gamma_j.$$

Accept the candidate draw with probability equal to the minimum of 1 and $\exp[-\alpha(\mu_* - \mu_1)]$.

6. Generate a candidate draw for μ_2, say μ_*, from an inverse gamma with shape parameter c and scale parameter d, where

$$c = 1 + K\nu_2 \qquad \text{and} \qquad d = \sum_{k=1}^{K} \nu_2 \tau_k.$$

Accept the candidate draw with probability equal to the minimum of 1 and $\exp[-\alpha(\mu_* - \mu_2)]$.

7. Generate a candidate draw for ν_1, say ν_*, according to

$$\nu_* = \nu_1 \exp(s),$$

where $s \sim N(0, t)$ and t, the precision of the Metropolis–Hastings proposal density, was chosen to be 1.1. Accept ν_* as the new value of ν_1 with probability equal to

$$\min\left(1, \frac{\nu_*^{-\delta} \exp(-\beta/\nu_*)}{\nu_1^{-\delta} \exp(-\beta/\nu_1)}\right.$$
$$\left. \times \prod_{j=1}^{J} \frac{(\gamma_j \nu_*/\mu_1)^{\nu_*} \exp(-\gamma_j \nu_*/\mu_1) \Gamma(\nu_1)}{(\gamma_j \nu_1/\mu_1)_1^{\nu} \exp(-\gamma_j \nu_1/\mu_1) \Gamma(\nu_*)}\right).$$

8. Generate a candidate draw for ν_2, say ν_*, according to

$$\nu_* = \nu_1 \exp(s),$$

where $s \sim N(0, t)$, and $t = .5$. Accept ν_* as the new value of ν_2 with probability equal to

$$\min\left(1, \frac{\nu_*^{-\delta} \exp(-\beta/\nu_*)}{\nu_2^{-\delta} \exp(-\beta/\nu_2)}\right.$$
$$\left. \times \prod_{k=1}^{K} \frac{(\tau_k \nu_*/\mu_2)^{\nu_*} \exp(-\tau_k \nu_*/\mu_2) \Gamma(\nu_2)}{(\tau_k \nu_2/\mu_2)_2^{\nu} \exp(-\tau_k \nu_2/\mu_2) \Gamma(\nu_*)}\right).$$

9. For $j = 1, \ldots, J$ and $i \in C_j$, generate candidate draws for $z_{i,j}$ from a truncated normal distribution with mean $\theta_i + \eta_{i, g(j)}$ and precision γ_j, truncated to the interval (c, d), where c is the largest value of $z_{k,j}$ for which $y_{k,j} < y_{i,j}$ and d is the smallest value of $z_{m,j}$ for which $y_{m,j} > y_{i,j}$. Accept the candidate draw as the new value of $z_{i,j}$ with probability equal to the minimum of 1 and the ratio of

$$\prod_{i \in C_j^*} p_{(i,j)}(\kappa),$$

evaluated at the candidate and the current values of $z_{i,j}$.

10. Generate a candidate draw for κ, κ_*, according to

$$\kappa_* = \kappa \exp(s),$$

where $s \sim N(0, t)$ and $t = 1.1$. Accept κ_* as the updated value of κ with probability equal to the minimum of 1 and the ratio

$$\frac{\kappa_*}{\kappa} \prod_{j=1}^{J} \prod_{i \in C_j^*} \frac{p_{(i,j)}(\kappa_*)}{p_{(i,j)}(\kappa)}.$$

The values of t chosen for the foregoing proposal densities resulted in acceptance rates in the range of 25%–35%.

4. ANALYSIS OF PRIMATE INTELLIGENCE DATA

The model described earlier for the analysis of correlated rank data was applied to the primate intelligence data reported in Table 1 by running the MCMC algorithm described in Section 3 for 40,000 burn-in iterations, followed by 4,000,000 parameter updates. The number of updates was determined using the coupling-regeneration scheme described by Johnson (1998), which indicated that for this total sample size, iterates separated by 40,000 updates in the chain could be considered effectively independent. More specifically, the probability that the distribution of 100 iterates separated by 40,000 updates in this MCMC chain would couple with the distribution of 100 independent draws from the posterior distribution exceeds .98. The coupling-regeneration scheme is probably conservative in this setting because of the difficulty associated with couplings of the $z_{i,j}$, which are drawn from truncated normal distributions with truncation points determined by other latent procedure variables. Nonetheless, the MCMC algorithm is relatively fast, requiring only 1.5 hours when executed on a 4-year-old Unix workstation.

As discussed in Section 3, the magnitudes of the paradigm–genus bias terms and the values τ, μ_2, and ν_2 were examined to assess support for paradigm–genus bias terms in the final model. The posterior means of the paradigm–genus precision parameters τ ranged from 36 to 65, whereas the posterior mean of μ_2 was 32. In comparison, the posterior means of the precision parameters γ ranged from 1.8 to 5.9, and the posterior mean of μ_1 was 4.6. Only two paradigm–genus bias estimates had posterior means exceeding .5 in magnitude, and neither could be reliably distinguished as being either positive or negative. Both effects occurred within the discrimination learning paradigm; the posterior mean of the discrimination learning–*Macaca* bias was $-.53$ with a posterior standard deviation of 1.11, and the posterior mean of the discrimination learning–*Presbytis* bias was $-.52$ with a standard deviation of 1.44. Note that *Presbytis* was rated only once and in that study was ranked highest. Mathematically, the fact that *Presbytis* received the highest rating from this study implies that the likelihood function does not provide a lower bound for the value of its latent procedure variables $z_{i,j}$. Thus a negative, but statistically insignificant, bias effect is not unexpected. Similar comments apply also to the discrimination learning–*Macaca* interaction. In five of the six procedures in this paradigm for which *Macaca* was ranked, it received the top rating. The posterior mean of each genera's latent intelligence variables based on the fit of the full model are provided in the first column of Figure 1.

The lack of significant bias terms suggests that a reduced model without paradigm–genus bias parameters is adequate for describing these data. Thus we refitted a reduced model using a simple modification of the MCMC algorithm described earlier; the proportion of times that each genus's global intelligence variable, θ_i, was greater than every other genus's intelligence variable was recorded, along with the MCMC sample mean and variance of these and related quantities. These proportions provide an estimate of the posterior odds that the general intelligence variable for one genus was greater than another, which in this particular study were the variables of
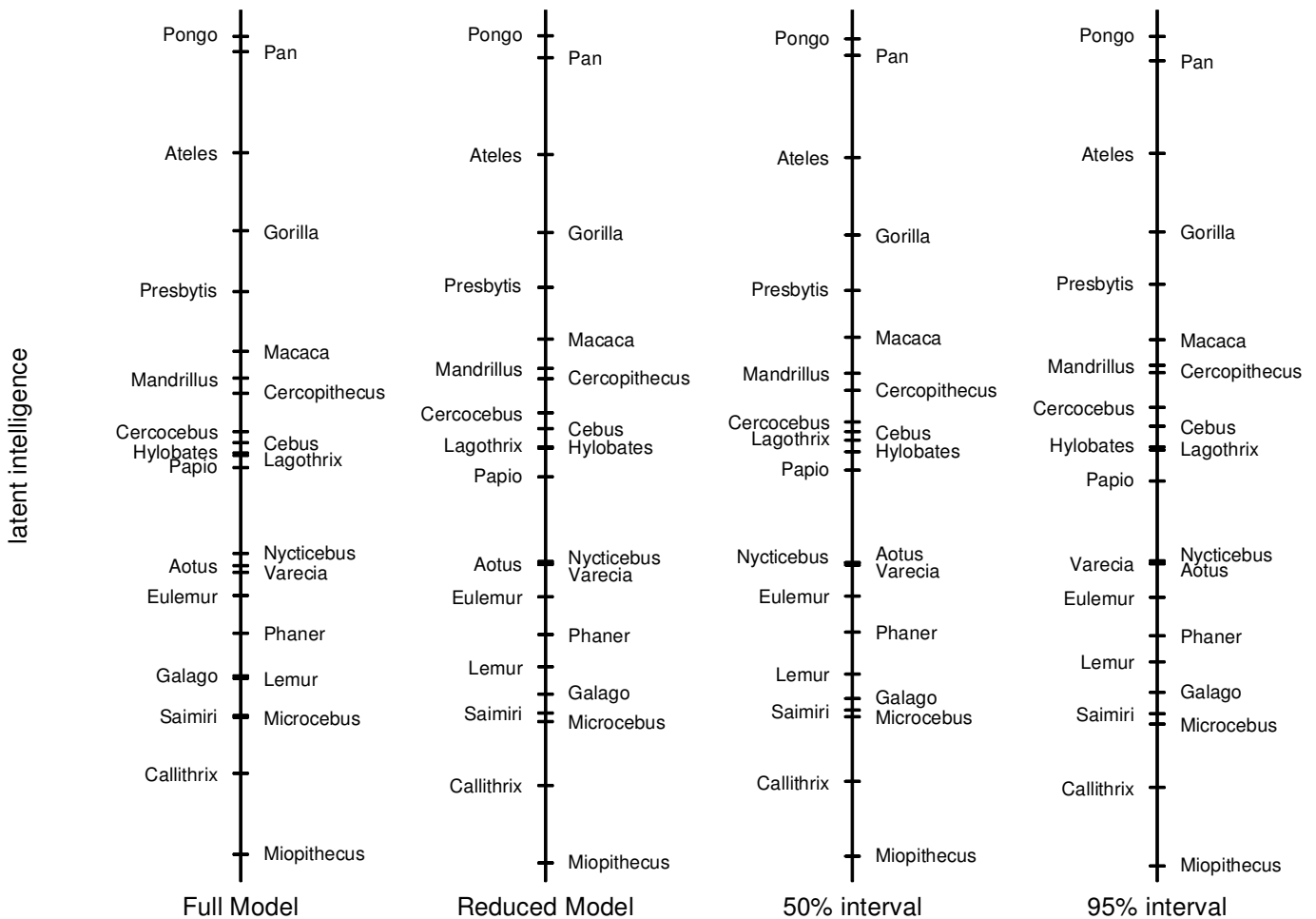
Figure 1.   *Posterior Means of Latent Intelligence Variables Under Varying Model Assumptions. The leftmost column depicts estimates obtained using the full model with paradigm–genus interactions. The next column shows the posterior means obtained under the reduced model for values of the hyperparameters speci'ed in Section 2. The third column displays the posterior means of the intelligence variables when the prior probability on ν_1 being in the interval (.25, 4) is .5. The 'nal column are the corresponding means when this prior probability is increased to .95.*

primary interest. These odds, as estimated from the reduced model, are displayed in Figure 2.

From a biological perspective, the odds displayed in Figure 2 confirm a widely-held belief that the great apes [*Gorilla*, *Pan* (chimpanzees), and *Pongo* (orangutan)] have greater cognitive ability than other nonhuman primates. The odds cited for *Ateles* are also of interest, because this genus is not closely related to the great apes, but nonetheless appears to rival *Gorilla* in terms of intelligence. *Presbytis* also is estimated to have high cognitive ability, but, because it was ranked in only one study, it had a comparatively large posterior variance. The posterior means of each genus's intelligence variable are displayed in the second column of Figure 1 for comparison with results obtained under the full model.

With the exception of procedures 15, 25, and 29, the posterior means of the procedure precision parameters ranged from 2.9 to 6.2. The precision parameters for these procedures were .9, .8, and 1.8. The comparatively low precision associated with procedures 15 and 29 can be understood through apparent discordancies in the rankings of the great apes in these procedures. *Macaca* is ranked higher than *Pan*, *Pongo*, and *Gorilla* in procedure 15, despite *Macaca's* lower rankings in relation to these genera in most other procedures. Similarly, procedure

29 yielded an unexpectedly low ranking of *Gorilla* and ranks *Papio* above *Pan*. In procedure 25, *Lemur* is ranked above *Cercopithecus*. Had similar discordancies been detected for other procedures in the same paradigms, it is likely that significant paradigm–genus biases would have been estimated, as we illustrate in Section 4.1. However, with only one discrepant procedure found within any paradigm, the model favors a slightly lower value of the relevant procedure precision parameters to a lower value of the paradigm precision parameters due to shrinkage effects imposed by the third-stage model.

The posterior mean of κ for this model was .26, with a 95% posterior probability interval extending from .13 to .47. At the posterior mean, this implies that latent procedure variables ($z_{i,j}$'s) separated by .17 units were estimated to have approximately a 50% chance of producing a tied ranking.

As a check on the sensitivity of these results to the choice of the third-stage hyperparameters α, β, and δ, histogram estimates of the parameters μ_1 and ν_1 were plotted against their corresponding priors in Figures 3 and 4. These figures suggest that the choice of the third-stage hyperparameters α, β, and δ probably weighed significantly in the determination of the posterior distributions on μ_1 and ν_1.
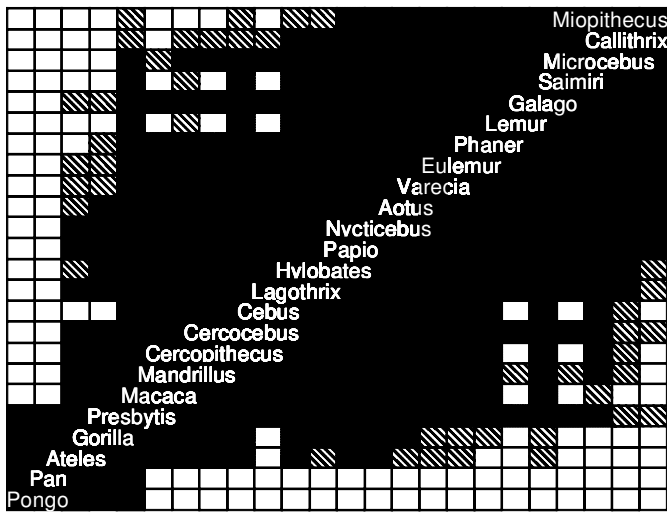
Figure 2. Posterior Probabilities that Latent Intelligence Variables of One Genera Exceeded the Latent Intelligence Variable of Another. Intervals describing these posterior probabilities can be found by observing the density of the plot immediately above or below the middle of a genus's name in the row corresponding to the genus for which a comparison is to be made. Solid regions indicate that the posterior odds that the leftmost genus was "smarter" than the rightmost genus were less than 9:1. The medium density regions indicate these posterior odds fell in the interval of 9:1 to 19:1. Low-density regions indicate the corresponding odds fell in the interval of 19:1 to 39:1. Blank regions indicate the posterior odds that one genus was smarter than the other exceeded 39:1. For example, the odds that Gorilla was "smarter" than Eulemur were estimated as being between 9:1 and 19:1.

To assess the sensitivity of the posterior distribution of θ to the choice of the priors on μ_1 and ν_1, we reestimated the model under differing prior assumptions regarding the values of α, β, and δ. Because the value of ν_1 controlled the amount of shrinkage of the procedure precision parameters toward their means, its value is important in determining the weights

given each of the procedures in estimating θ. Sensitivity of the posterior on θ to ν_1 is thus of special interest; for that reason, we have displayed posterior means of θ under values of β and δ that placed 95% and 50% of the prior mass of ν_1 on the interval (.25, 4). (The baseline model assigned 80% mass to this interval.) The resulting posterior means are displayed, along with the results from the full model with paradigm–genus biases and the reduced, baseline model, in Figure 1.

Values of the posterior means of the cognitive ability variables in Figure 1 indicate that although the posterior distribution on ν_1 is sensitive to the values of the third-stage hyperparameter β and δ, the posterior distributions on the primary variables of interest, θ, are relatively insensitive to the choice of this hyperparameter.

As a cursory check of model fit, 500 replications of the ranking reported in Table 1 were generated from the posterior distribution. For each of these posterior-predictive replications, the proportion of comparisons within each procedure that were correctly ranked, according to the posterior mean of θ, were computed. These proportions are compared to the proportion of the observed comparisons that were correctly predicted by the posterior mean of θ in Figure 5. The proportion of ranks correctly predicted by the posterior means of the cognitive ability variables was 85% for the observed data. This value fell in the middle of the range of proportions correctly predicted for data simulated from the posterior, thus providing some evidence of model adequacy.

4.1 Detection of Paradigm–Genus Interactions

A thorough investigation of the power of the proposed model in detecting paradigm–genus biases is not possible here because of space constraints. However, a simple illustration that is suggestive of the model's performance in this regard can be constructed by creating an artificial paradigm that contains probable paradigm–genus biases. To this end, the data from
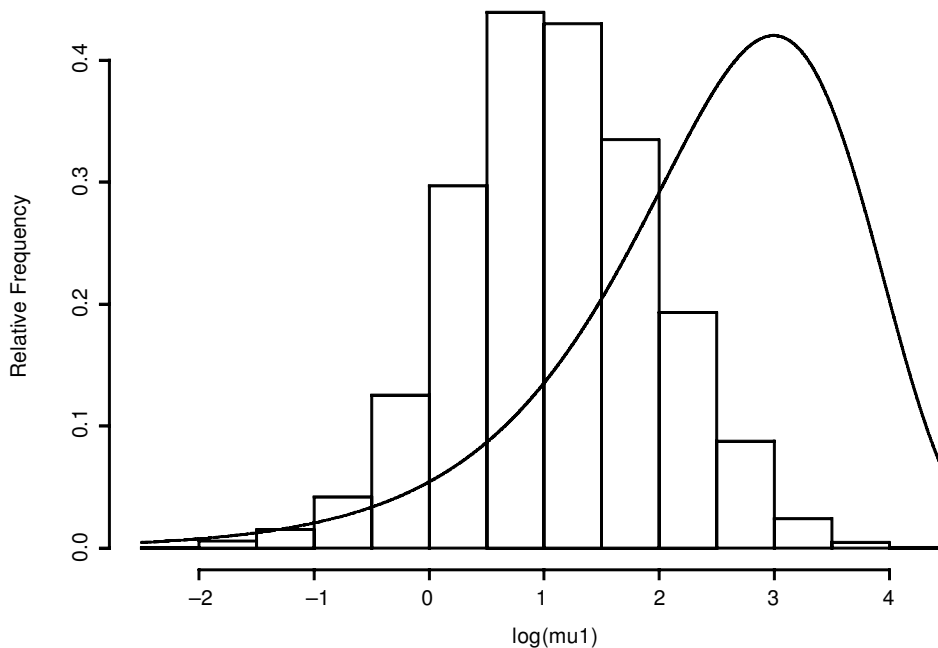

Figure 3. Histogram Estimate of the Posterior Distribution of log(μ_1). The (rescaled) prior density is depicted as a solid line for comparison.
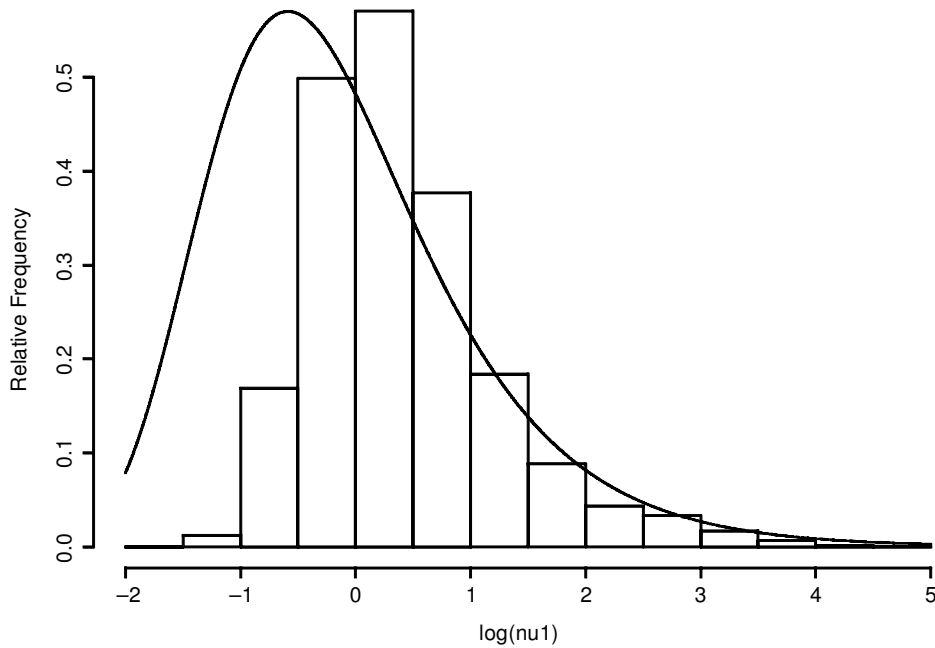
Figure 4. *Histogram Estimate of the Posterior Distribution of log(ν_1). The (rescaled) prior density is depicted as a solid line for comparison.*

the 15th procedure, which exhibited a small precision and had only five rankings, was artificially replicated a second and third time and appended to the original data as a tenth paradigm. Procedure 15 was also chosen because it was the most interesting procedure from a substantive perspective; the great apes performed uniformly worse than *Macaca* in this procedure and were not tested in any other procedures in this paradigm.

Refitting the full model with the added paradigm had little effect on the latent cognitive parameters $\boldsymbol{\theta}$. However, the pos-

terior median of the precision parameter of the tenth paradigm was estimated to be .04, whereas the posterior medians of the first nine paradigm precision parameters ranged from 13 to 37. In addition, four of the five paradigm–genus interaction terms estimated for this paradigm had posterior means larger than 1 in magnitude. The interaction of the tenth paradigm with *Hylobates* had a posterior mean of 6.5 and a posterior standard deviation of 5.0; that with *Macaca*, a posterior mean of −7.8 and standard deviation of 4.7; that with *Pan*, a mean
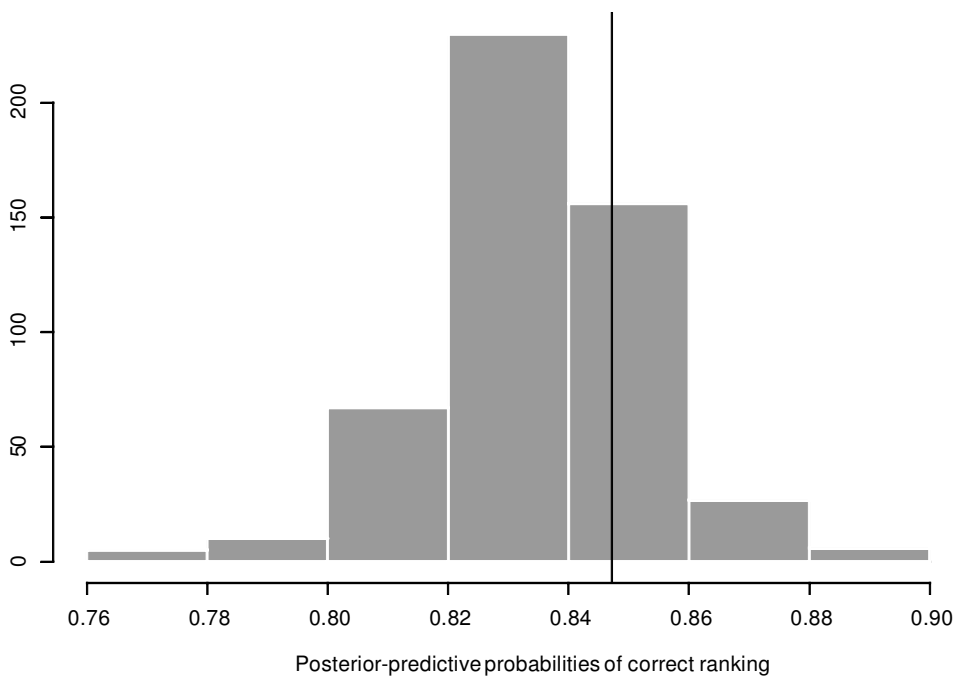

Figure 5. *Histogram Estimate of the Probabilities That Pairwise Comparisons of Genera Match the Ranking Predicted by the Posterior Mean of $\boldsymbol{\theta}$, for Replications of the Data Generated From the Predictive Distribution. The vertical line depicts the proportion of comparisons correctly predicted for the observed data.*

of −2.8 and a standard deviation of 3.7; and that with *Pongo*, a mean of 4.0 and a standard deviation of 3.9. The model thus provides a clear indication of either paradigm–genus biases for the artificial paradigm or, under an alternative interpretation, the presence of a higher-dimensional trait variable.

5. DISCUSSION

The question of whether some primate taxa outperform others across a range of experimental paradigms has been of long-standing interest to those concerned with the organization and evolution of cognitive abilities. Unfortunately, before the development of the proposed model, statistical methods that could simultaneously accommodate global trait variables, paradigm–genus biases, tied ranks, and missing data were unavailable, making the combined analysis of relevant studies impossible. Our findings that paradigm–genus biases are insubstantial and that many genera differ significantly from others on their global trait variables indicates that, contrary to much current thinking (e.g., Tooby and Cosmides 1992; Tomasello and Call 1997; Shettleworth 1998), the intuitive idea that taxa differ in domain-general ability or intelligence might yet be vindicated.

Of course, the fact that we did not detect paradigm–genus bias effects does not mean that we would not detect them if more data were available. Nevertheless, the global trait variables estimated without these effects correctly predict performance in 194 of 229 genus-by-genus comparisons (85%), exactly the same proportion predicted by the full model containing paradigm–genus bias effects. This statistic underscores the fact that even if more data were included, the explanatory power of paradigm–genus effects would still be small relative to the global trait variable.

From a statistical perspective, the proposed model for the analysis of multistudy rank data provides several innovations over extant Thurstonian models. Modeling ties through the introduction of the parameter κ greatly simplifies estimation of latent trait values using information derived from tied ranks. The simplicity of this mechanism contrasts sharply with classical Thurstonian models in which ties are not easily accommodated (see, e.g., Marden 1995). For datasets like the nonhuman primate intelligence meta-analysis, this aspect of the model is particularly important, because ties comprise approximately one-third of the data values.

In addition, the Bayesian formulation of the model simplifies the assessment of uncertainty. Because in practice individual procedures or raters usually do not rank more than 15 or so items, the posterior distribution on model parameters is especially helpful in this regard. Finally, the hierarchical structure imposed on the precision parameters allows the model to "borrow strength" in estimating precision parameters for studies in which only limited data are obtained. Again, because rank data are generally not extensive, this too is an important feature of the model.

REFERENCES

Balda, R. P., Kamil, A. C., and Bednekoff, P. A. (1996), "Predicting Cognitive Capacity from Natural History: Examples from Four Species of Corvids," in *Current Ornithology*, eds. V. Nolan, Jr. and E. D. Ketterson, New York: Plenum Press, pp. 33–36.

Balasch, J., Sabater-Pi, J., and Padrosa, T. (1974), "Perceptual Learning Ability in *Mandrillus sphinx* and *Cercopithecus nictitans*," *Revista Española de Fisiologia*, 30, 15–20.

Böckenholt, U. (1992), "Thurstonian Representation of Ranking Data," *British Journal of Mathematical and Statistical Psychology*, 45, 31–50.

—— (1993), "Applications of Thurstonian Models to Ranking Data," in *Probability Models and Statistical Analyses for Ranking Data*, eds. M. A. Fligner and J. S. Verducci, New York: Springer-Verlag, pp. 157–172.

Bradley, R. A. (1984), "Paired Comparisons: Some Basic Procedures and Examples," in *Handbook of Statistics 4*, eds. P. R. Krishnaiah and P. K. Sen, Amsterdam: Elsevier, pp. 299–326.

Bradley, R. A., and Terry, M. E. (1952), "The Rank Analysis of Incomplete Block Designs: 1, The Method of Paired Comparisons," *Biometrika*, 39, 324–345.

Cotterman, T. E., Meyer, D. R., and Wickens, D. D. (1956), "Discrimination Reversal Learning in Marmosets," *Journal of Comparative and Physiological Psychology*, 49, 539–541.

Crawford, F. T. (1962), "Reversal Learning to Spatial Cues by Monkeys," *Journal of Comparative and Physiological Psychology*, 55, 869–871.

Daniels, H. E. (1950), "Rank Correlation and Population Models," *Journal of the Royal Statistical Society*, Ser. B, 12, 171–181.

David, H. A. (1988), *The Method of Paired Comparisons* (2nd ed.), London: Chapman and Hall.

Davidson, R. R. (1970), "On Extending the Bradley–Terry Model to Accommodate Ties in Paired-Comparison Experiments," *Journal of the American Statistical Association*, 65, 317–328.

Davis, R. T., and Leary, R. W. (1968), "Learning of Detour Problems by Lemurs and Seven Species of Monkeys," *Perceptual and Motor Skills*, 27, 1031–1034.

Davis, R. T., Leary, R. W., Stevens, D. A., and Thompson, R. F. (1967), "Learning and Perception of Oddity Problems by Lemurs and Seven Species of Monkey, *Primates*, 8, 311–322.

Davis, R. T., McDowell, A. A., and Nissen, H. W. (1957), "Solution of Bent-Wire Problems by Monkeys and Chimpanzees," *Journal of Comparative and Physiological Psychology*, 49, 20–25.

Deaner, R. O., van Schaik, C., and Johnson, V. E. (2001), "Do Some Species Have Better Domain-General Cognition Than Others? A Meta-Analysis of Non-Human Primate Studies of the 20th Century," unpublished manuscript.

de Blois, S. T., Novak, M. A., and Bond, M. (1998), "Object Permanence in Orangutans (*Pongo pygmaeus*) and Squirrel Monkeys (*Saimiri sciureus*)," *Journal of Comparative Psychology*, 112, 137–152.

—— (1999), "Can Memory Requirements Account for Species' Differences in Invisible Displacement Tasks?," *Journal of Experimental Psychology*, 25, 168–176.

Finch, G. (1941), "The Solution of Patterned String Problems by Chimpanzees," *Journal of Comparative Psychology*, 32, 83–90.

Fischer, G. J., and Kitchener, S. L. (1965), "Comparative Learning in Young Gorillas and Orang-utans," *Journal of Genetic Psychology*, 107, 337–348.

Garcha, H. S., and Ettlinger, G. (1979), "Object Sorting by Chimpanzees and Monkeys," *Cortex*, 15, 213–224.

Glickman, M. E. (1999), "Parameter Estimation in Large Dynamic Paired Comparison Experiments," *Applied Statistics*, 48, 377–394.

Glickman, M. E., and Stern, H. S. (1998), "A State-Space Model for National Football League Scores," *Journal of the American Statistical Association*, 93, 25–35.

Gossette, R. L. (1970), "Comparisons of SDR Performance of Gibbons and Three Species of New World Monkeys on a Spatial Task," *Psychonic Science*, 19, 301–303.

Gossette, R. L., and Inman, N. (1966), "Comparison of Spatial Successive Discrimination Reversal Performances of Two Groups of New World Monkeys," *Perceptual and Motor Skills*, 23, 169–170.

Gossette, R. L., and Slonim, R. L. (1969), "Spatial SDR Performance Across Three Species of New World Monkeys," *Psychonic Science*, 14, 222–223.

Harlow, H. F., and Settlage, P. H. (1934), "Comparative Behavior of Primates VII. Capacity of Monkeys to Solve Patterned String Tests," *Journal of Comparative Psychology*, 18, 423–435.

Harlow, H. F., Uehling, H., and Maslow, A. H. (1932), "Comparative Behavior of Primates. I. Delayed Reaction Tests on Primates from the Lemur to the Orang-utan," *Journal of Comparative Psychology*, 13, 313–343.

Johnson, V. E. (1996), "On Bayesian Analysis of Multirater Ordinal Data," *Journal of the American Statistical Association*, 91, 42–51.

—— (1997), "An Alternative to Traditional GPA for Evaluating Student Performance," *Statistical Science*, 12, 251–278.

—— (1998), "A Coupling-Regeneration Scheme for Diagnosing Convergence in Markov Chain Monte Carlo Algorithms," *Journal of the American Statistical Association*, 93, 238–248.

Johnson, V. E., and Albert, J. H. (1999), *Ordinal Data Modeling*, Springer-Verlag: New York.

Lefebvre, L., and Giraldeau L.-A. (1996), "Is Social Learning an Adaptive Specialization?," in *Social Learning in Animals: The Roots of Culture*, eds. C. M. Heyes and B. G. Galef Jr., London: Academic Press, pp. 107–128.

Manocha, S. N. (1967), "Discrimination Learning in Langurs and Rhesus Monkeys," *Perceptual and Motor Skills*, 24, 805–806.

Marden, J. I. (1995), *Analyzing and Modeling Rank Data,* London: Chapman and Hall.

Maslow, A. H., and Harlow, H. F. (1932), "Comparative Behavior of Primates," II. Delayed Reaction Tests on Primates at the Bronx Park Zoo," *Journal of Comparative Psychology*, 14, 97–107.

McDowell, A. A., and Nissen, H. W. (1959), "Solution of a Bi-Manual Coordination Problem by Monkeys and Chimpanzees," *Journal of Genetic Psychology*, 94, 35–42.

Meyer, D. R. (1951), "Food Deprivation and Discrimination Reversal Learning by Monkeys," *Journal of Experimental Psychology*, 41, 10–16.

Miles, R. C. (1957a), "Delayed-Response Learning in the Marmoset and the Macaque," *Journal of Comparative Physiological Psychology*, 50, 352–355.

—— (1957b), "Learning-Set Formation in the Squirrel Monkey," *Journal of Comparative Psychology*, 50, 356–357.

Miles, R. C., and Meyer, D. R. (1956), "Learning Sets in Marmosets," *Journal of Comparative Psychology*, 49, 219–222.

Natale, F., and Antinucci, F. (1989a), "Stage 6 Object-Concept Representation," in *Cognitive Structure and Development in Nonhuman Primates*, ed. F. Antinucci, Hillsdale, NJ: Lawrence Erlbaum, pp. 97–112.

—— (1989b), "Causality II: The Stick Problem," in *Cognitive Structure and Development in Nonhuman Primates*, ed. F. Antinucci, Hillsdale, NJ: Lawrence Erlbaum, pp. 121–133.

Natale, F., Antinucci, F., Spinozzi, G., and Poti, P. (1986), "Stage 6 Object Concept in Non-Human Primate Cognition: A Comparison between Gorilla (*Gorilla gorilla gorilla*) and Japanese Macaque (*Macaca fuscata*)," *Journal of Comparative Psychology*, 100, 335–339.

Ohta, H. (1983), "Learning Set Formation in Slow Lorises (*Nycticebus coucang*)," *Folia Primatologia*, 40, 256–267.

Ohta, H., Ishida, H., and Matano, S. (1984), "Learning Set Formation in Ring-Tailed Lemurs (*Lemur catta*)," *Folia Primatologia*, 43, 53–58.

—— (1987), "Learning Set Formation in Thick-Tailed Bush Babies (*Galago crassicaudatus*) and Comparison of Learning Ability Among Four Species," *Folia Primatologia*, 48, 1–8.

Riesen, A. H., Greenberg, B., Granston, A. S., and Fantz, R. L. (1953), "Solutions of Patterned String Problems by Young Gorillas," *Journal of Comparative Psychology*, 46, 19–22.

Rumbaugh, D. M. (1997), "Competence, Cortex, and Primate Models: A Comparative Primate Perspective," in *Development of the Prefrontal Cortex*, eds. N. A. Krasnegor, G. R. Lyon, and P. S. Goldman-Rakic, Baltimore: P. H. Brooks Publishing Company, pp. 117–139.

Rumbaugh, D. M., and Arnold, R. C. (1971), "Learning: A Comparative Study of Lemur and Cercopithecus," *Folia Primatologica*, 14, 154–160.

Rumbaugh, D. M., and McCormack, C. (1967), "The Learning Skills of Primates: A Comparative Study of Apes and Monkeys," in *Progress in Primatology*, eds. D. Starck, R. Schneider, and H. J. Ruhn, Stuttgart: Gustav Fischer Verlag, pp. 289–306.

Rumbaugh, D. M., Sammons, M. E., Prim, M. M., and Philips, S. (1965), "Learning Set in Squirrel Monkeys as Affected by Pretraining with Differentially Rewarded Single Objects," *Perceptual and Motor Skills*, 21, 63–70.

Rumbaugh, D. M., Savage-Rumbaugh, E. S., and Washburn, D. A. (1996), "Toward a New Outlook on Primate Learning and Behavior: Complex Learning and Emergent Processes in Comparative Perspective," *Japanese Psychological Research*, 38, 113–125.

Schrier, A. M. (1972), "Learning-Set Formation and Transfer in Rhesus and Talapoin Monkeys," *Folia Primatologia*, 17, 389–396.

Shell, W. F., and Riopelle, A. J. (1958), "Progressive Discrimination Learning in Platyrrine Monkeys," *Journal of Comparative Psychology*, 51, 467–470.

Shettleworth, S. J. (1998), *Cognition, Evolution, and Behavior*, Oxford, U.K.: Oxford University Press.

Stern, H. (1990), "Models for Distributions on Permutations," *Journal of the American Statistical Association*, 85, 558–564.

Stevens, D. A. (1965), "A Comparison of Learning in Rhesus Monkeys, Cebus Monkeys, Lemurs, and Burmese Cats," unpublished doctoral dissertation, University of Oregon.

Strong, P. N. J., and Hedges, M. (1966), "Comparative Studies in Simple Oddity Learning: I. Cats, Raccoons, Monkeys, and Chimpanzees," *Psychonomic Science*, 5, 13–14.

Thomas, R. K., and Boyd, M. G. (1973), "A Comparison of *Cebus albifrons* and *Saimiri sciureus* on Oddity Performance," *Animal Learning and Behavior*, 1, 151–153.

Thurstone, L. L. (1927), "A Law of Comparative Judgement," *Psychological Review*, 15, 284–297.

—— (1931), "Rank Order as a Psychological Method," *Journal of Experimental Psychology,*" 14, 187–201.

Tinklepaugh, O. L. (1932), "Multiple Delayed Reaction With Chimpanzees and Monkeys," *Journal of Comparative Psychology*, 13, 207–243.

Tomasello, M., and J. Call, (1997), *Primate Cognition*, Oxford, U.K.: Oxford University Press.

Tooby, J., and Cosmides, L. (1992), "The Psychological Foundations of Culture," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, eds. J. H. Barkow, L. Cosmides, and J. Tooby, Oxford, U.K.: Oxford University Press, pp. 19–136.

Visalberghi, E., Rumbaugh, S. S., and Fragaszy, D. M. (1995), "Performance in Tool Using Task by Common Chimpanzees (*Pan troglodytes*), Bonobos (*Pan paniscus*), an Orangutan (*Pongo pygmaeus*), and Capuchin Monkeys (*Cebus apella*)," *Journal of Comparative Psychology*, 109, 52–60.