Bayesian analysis of linear dominance hierarchies

ELDRIDGE S. ADAMS

Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs

Studies on social animals often seek to identify dominance hierarchies, in which individuals are ranked according to competitive abilities based on counts of wins and losses in pairwise encounters. I illustrate Bayesian approaches, based on the method of paired comparisons, for determining ranks and for estimating relationships between dominance ability and other attributes. Bayesian inference combines prior probability distributions for each unknown parameter with likelihood functions to produce the joint posterior probability distribution for the quantities of interest. In contrast to nonparametric techniques for inferring ranks, Bayesian models yield measures of certainty for each inference and allow rigorous estimates of correlations between ranks and covariates even when there is considerable uncertainty as to the ranks themselves. A possible objection to the Bayesian approach is that it appears to entail more restrictive assumptions than do simpler methods. However, simulations show that Bayesian inferences are more robust to deviations from these assumptions than are the results of nonparametric methods.

© 2005 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Social groups are often characterized by dominance hierarchies (Drews 1992). The concept of a dominance hierarchy implies that individuals differ in ability to prevail in a particular type of competition and that it is meaningful to rank these abilities from highest to lowest. Data are commonly summarized in matrix form, indicating the number of contests won by each competitor in contests with each other competitor (e.g. Table 1). Statistical analyses often seek to identify the rank order of competitive abilities and to estimate relationships between ranks and covariates, such as body size or mating success. This paper describes how these analyses can be conducted within a Bayesian framework, and evaluates the advantages and disadvantages of this approach.

Most studies on animal hierarchies rely on nonparametric methods to infer ranks. The current standard among nonparametric models is the I&SI method, an abbreviation for 'inconsistencies and strengths of inconsistencies' (de Vries 1998; de Vries & Appleby 2000). An inconsistency is any pair for which the individual assigned the higher rank lost, or showed subordinate behaviour, in a majority of encounters. The number of ranks separating these individuals is the strength of the inconsistency. The I&SI method seeks the ordering with the minimum number of inconsistencies and, to break

ties, the minimum total strength of inconsistencies. While nonparametric methods provide common-sense procedures for deciding among possible orderings, they yield little information about the degree of certainty that one can assign to the results. As I will show, for many typical data sets the available methods are not likely to determine the true rank order of a set of competitors unambiguously. It is therefore desirable to have some measure of the degree of credibility for a particular result, to seek more limited claims about which one can be more certain, and to retain information on uncertainty in any further statistical analyses.

To measure associations between dominance rank and other attributes of the animals, a typical approach is to accept the ordering suggested by nonparametric or maximum likelihood analysis as being true, and then to measure the rank correlation with another measured quantity. This approach ignores uncertainty in ranks, and so can produce misleading inferences about the causes or consequences of dominance. By contrast, Bayesian analyses produce probability distributions for each inference, quantifying the degree of belief that can be assigned to possible parameter values (Gill 2002; Gelman et al. 2004). By expressing results as probability distributions, information about uncertainty in dominance structure can be summarized and carried into further analyses. This is particularly important when sample sizes are small and uneven.

Methods based on explicit probability models have also been developed (e.g. Leonard 1977; Boyd & Silk 1983;

Correspondence: E. S. Adams, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269-3043, U.S.A. (email: eldridge.adams@uconn.edu).

Table 1. Matrix of wins and losses for five cockroaches, *Nauphoeta cinerea* (Bell & Gorton 1978)

			Loser		
	A	B	C	D	E
Winner					
A	—	9	12	6	27
B	10	—	9	12	12
C	2	5	—	0	2
D	3	3	0	—	2
E	2	3	0	4	—

Tufto et al. 1998) but are rarely used. Most are based on the method of paired comparisons (David 1988), which is readily interpreted as a method of determining relative dominance abilities based on counts of wins and losses. Boyd & Silk (1983) presented methods for producing maximum likelihood estimates of dominance indexes under the Bradley–Terry model, a version of the method of paired comparisons (Bradley & Terry 1952; David 1988). The Bradley–Terry model assumes the existence of a transitive linear hierarchy for which the probability per encounter that one individual prevails over another is a logistic function of the difference in dominance indexes between these two animals (Fig. 1). Tufto et al. (1998) described a method, also based on the Bradley–Terry model, for evaluating whether the assumption of transitivity is warranted.

This paper describes Bayesian versions of the method of paired comparisons and their advantages for behavioural studies, extending the work of Davidson & Solomon (1973) and of Leonard (1977). These previous contributions have been ignored by behavioural biologists, presumably due to lack of familiarity with Bayesian methods. Recently, however, Bayesian approaches have become more common in fields related to animal behaviour (e.g. Hilborn & Mangel 1997; Holsinger et al. 2002; Holder &

Lewis 2003). Also, development of software implementing Markov chain Monte Carlo techniques has made Bayesian analysis more accessible (Gill 2002). Here, I show how Bayesian hierachy analysis can be carried out with the WinBUGS software package (Spiegelhalter et al. 2003), how posterior probabilities can be assigned to possible rank orderings, and how regressions of dominance ability on covariates can be estimated. I then address a possible cause of resistance to use of the method of paired comparisons; namely, its apparent reliance on assumptions that may be violated in animal studies (de Vries 1998). By simulating data under a variety of assumptions, including cases in which the assumptions of the method of paired comparisons are violated, I show that nonparametric methods for identifying ranks are sensitive to the same assumptions, while yielding less information and reduced accuracy.

ADVANTAGES OF BAYESIAN APPROACHES

Methods based on probability models offer a number of benefits relative to nonparametric techniques. In addition to inferring which estimate of ranks has the greatest support, probability-based models yield measures of certainty about the inferences, such as 95% confidence intervals. They also allow analysis of deviations between model assumptions and data, and comparison of alternative models, such as those assuming transitive or intransitive dominance relationships (Tufto et al. 1998). In addition, the method of paired comparisons produces estimates of dominance abilities on an interval scale, so that the magnitude of differences can be examined as well as the ranks (Boyd & Silk 1983). This in turn allows use of parametric regression models with dominance ability as either a dependent or independent variable.

Among methods based on probability models, Bayesian inference offers several advantages over maximum likelihood estimates. The first is that the results allow straightforward statements about the probability that a hypothesis is true or that a parameter lies within a particular range (Gill 2002; Congdon 2003). For example, the method described here yields a posterior probability for each possible ordering of individuals, which is the probability that that ordering is correct, given the model and the data. A Bayesian 95% credibility interval, unlike a traditional confidence interval, is properly interpreted as containing the true value of the parameter with a probability equal to 0.95, given the model and the data (Gill 2002).

A second advantage of Bayesian methods is that prior information can be incorporated into the analysis. For the analysis of dominance hierarchies, the inclusion of prior information leads to an important pragmatic advantage; namely, that the resulting estimates of competitive abilities are always finite (Leonard 1977). By contrast, algorithms used to produce maximum likelihood estimates of dominance abilities under the Bradley–Terry model fail to converge to finite values, and so cannot be used, for many data sets. This is explained further below. Finite estimates of competitive ability offer greater biological realism as
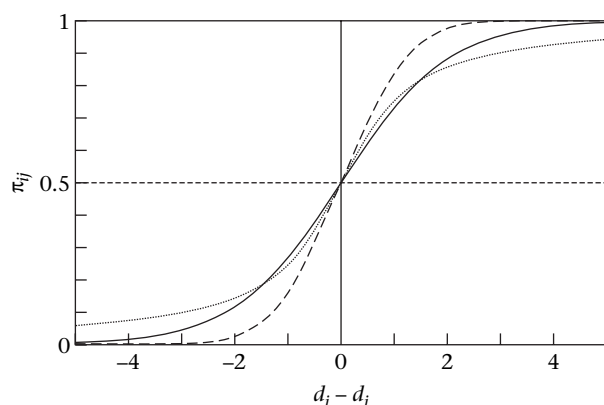

Figure 1. Alternative functions describing π_{ij}, the probability that individual A_i wins in an encounter with individual A_j, as a function of the difference between their dominance indexes, $d_i - d_j$. The Bradley–Terry model assumes a logit link (solid curve). Other possibilities include the probit link (dashed curve) and the cumulative Cauchy distribution (dotted curve).

well as computational advantages, particularly in regression models.

A third advantage is the comparative ease with which various sources of uncertainty can be incorporated accurately into the analysis. Bayesian methods estimate the probability distribution of parameters under complex models without relying on large-sample approximations to normality (Congdon 2003).

BAYESIAN ANALYSIS OF RANK ORDER

The Method of Paired Comparisons

The method of paired comparisons (David 1988) can be interpreted as a method to reveal underlying dominance abilities given data on the outcome of contests (Leonard 1977; Boyd & Silk 1983). 'Dominance ability' is understood here as a measure of ability to prevail in a particular type of encounter due to fighting ability, motivation and past experiences. Physical strength and learning through fighting and observation are among the possible determinants of dominance ability.

Let d_i represent the dominance index of the ith individual A_i. Let n_{ij} represent the number of encounters observed between a pair of individuals, A_i and A_j. The number of wins, W_{ij}, by individual A_i in encounters with A_j is assumed to be binomially distributed conditional on n_{ij} and probability $\pi_{ij} = H(d_i - d_j)$, where H is a function of the difference between the dominance indexes of A_i and A_j. The function H, called the link function, maps differences in strength, for which the scale ranges from negative infinity to positive infinity, to a probability that must lie between 0 and 1 inclusive (Fig. 1). Possibilities include the logit link, as assumed by the Bradley–Terry model, and the probit link, among others (Fig. 1). For reasonable link functions, $\pi_{ij} = 0.5$ when $d_i = d_j$; that is, that each individual has a 50% chance of prevailing when the dominance abilities of the two individuals are equal (see Fig. 1)

Where two individuals meet in n_{ij} encounters, the probability that the random variable W_{ji} equals w_{ij}, the observed number of wins by individual A_i, is

$$\binom{n_{ij}}{w_{ij}} \pi_{ij}^{w_{ij}} \pi_{ji}^{w_{ji}} \tag{1}$$

Since the value of $\binom{n_{ij}}{w_{ij}}$ does not depend on the d_i values, the likelihood of the data for all pairs of individuals, given the d_i is

$$\prod_{i<j} \pi_{ij}^{w_{ij}} \pi_{ji}^{w_{ji}} = \prod_{i \neq j} \pi_{ij}^{w_{ij}} = \prod_{i \neq j} H(d_i - d_j)^{w_{ij}} \tag{2}$$

Under the Bradley–Terry model, $H(d_i - d_j) = 1/(1 + \exp(-(d_i - d_j)))$.

Bayesian Estimation

Bayesian inference combines (1) a likelihood function with (2) prior probability distributions for each model parameter to produce (3) posterior probability distributions for quantities of interest (Congdon 2003; Gelman et al. 2004). In essence, Bayesian analysis uses the data to move from a state of great uncertainty about parameter values, described by prior probability distributions, to a state of greater certainty, described by posterior probability distributions.

(1) The likelihood function quantifies the probability of observing the data given particular parameter values relative to the probability of the data under other parameter values. For the method of paired comparisons, the likelihood is proportional to the right hand side of equation 2.

(2) Prior probability distributions, or priors, describe beliefs or information about parameter values prior to the observations that form the data set. Newcomers to Bayesian methods are often uncomfortable about specifying priors since this requires subjective decisions when little hard information is available for guidance. The solution is to use diffuse or 'vague' priors, for which plausible parameter values are nearly equally probable, and to test the sensitivity of the inferences to the choice of priors post facto. If there is no prior information about the animals' dominance abilities, it is reasonable to use the same symmetrical prior probability distribution, centred on the same value, for each individual's dominance index. A normal prior distribution with a large variance corresponds to the prior belief that modest differences in dominance abilities are nearly equally likely, that very large differences are somewhat less probable, and that no animal's dominance ability is infinitely higher than that of any other animal.

Even with vague prior probability distributions, Bayesian estimation solves a problem often encountered in maximum likelihood estimation of the Bradley–Terry model; specifically, that maximum likelihood estimates of differences in dominance ability may be infinite (Leonard 1977). This problem arises if the individuals can be partitioned such that members of one subgroup won all encounters with members of the other subgroup (Boyd & Silk 1983), a common occurrence in studies on animal hierarchies. For example, the top two individuals in Table 2 never lost and the last individual never won in the observed contests; therefore, the maximum likelihood estimates of dominance abilities are infinitely greater for the top two individuals and infinitely lower for the bottom individual than they are for the other animals. Given that animals have finite strengths and imperfect perceptual abilities, it is unreasonable to infer that one animal's competitive ability is infinitely higher or lower than that of another animal. Bayesian priors can prevent biologically implausible estimates while also conferring computational advantages.

(3) Posterior probability distributions describe the degree of belief that can be assigned to possible values of the model parameters, summarizing information contained in the data and the prior distributions. As the sample size increases, the influence of the priors on the posteriors diminishes. The interpretation of posterior probabilities is straightforward, assuming that an appropriate model has been used. For example, if 99% of the posterior

Table 2. Matrix of wins and losses for 20 female bighorn sheep, *Ovis canadensis*, and their ages (Hass 1991)

	Loser																				Age
	15	17	21	06	08	01	10	22	02	11	05	16	12	09	26	25	29	27	28	24	
Winner																					
15	—	0	4	0	7	0	0	1	1	0	3	0	4	1	1	2	2	9	2	5	7+
17	0	—	3	0	1	0	0	0	0	1	2	0	5	1	0	0	0	2	1	1	7+
21	0	0	—	0	2	0	1	1	1	1	1	1	1	1	3	0	1	3	2	1	+
06	0	0	0	—	0	0	0	1	0	0	0	0	3	0	1	0	0	1	1	1	5
08	0	0	0	1	—	1	0	2	3	2	1	0	1	1	2	2	2	6	4	0	6
01	0	0	0	0	0	—	0	0	0	0	0	0	0	0	1	0	1	0	0	0	6
10	0	0	0	0	0	0	—	0	0	0	2	0	0	0	0	0	0	1	0	0	4
22	0	0	0	0	0	0	0	—	2	1	2	0	1	0	2	0	0	1	2	3	6
02	0	0	0	0	0	0	0	0	—	0	1	2	1	4	1	0	0	1	0	0	3
11	0	0	0	0	1	0	0	0	0	—	0	0	1	0	1	0	0	1	0	1	5
05	0	0	0	0	0	0	0	0	0	0	—	0	0	0	0	1	1	2	1	1	2
16	0	0	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	1	0	5
12	0	0	0	0	0	0	0	0	0	0	5	0	—	0	0	1	1	1	0	0	2
09	0	0	0	0	0	0	0	0	0	0	0	0	0	—	0	1	0	0	1	0	2
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	1	1	0	1	0	2
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	0	0	1	0	1
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	0	1	0	1
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	0	1	1
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	3	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	1

distribution for an individual's dominance ability lies between 0.2 and 0.8, then the true value lies within that range with a probability of 0.99.

Implementation of Bayesian Models

Bayesian models were implemented in WinBUGS 1.4 (Spiegelhalter et al. 2003). WinBUGS, which is currently distributed free of charge, allows estimation of a wide variety of Bayesian models by use of Markov chain Monte Carlo (MCMC) simulation. The user's manual demonstrates how to specify a model, how to use the software to estimate parameters, and how to interpret the output. MCMC methods approximate posterior distributions by drawing large numbers of samples from them. By increasing the lengths of the Markov chains, summaries of the distribution, such as means, can be estimated to an arbitrary degree of precision. The appendices list the WinBUGS model specifications for the examples in this paper.

While the computations are handled by the software, the user must be careful to ensure that the Markov chains are of sufficient length to converge to the posterior distribution and to sample it thoroughly (see Gill 2002). Because the initial values of the Markov chain are usually chosen arbitrarily, the early steps of the chain may draw samples from regions of low posterior probability. Therefore, the earlier samples, referred to as a burn-in or dememorization period (Gill 2002) are discarded. WinBUGS supports diagnostics that help to determine how many early samples to discard, and how long the chain should run thereafter. For the analyses in this paper, I ran two or more chains from overdispersed initial values, examined trace plots and quantile plots by eye to determine the length of the discarded burn-in, and to ensure

that means and 95% credibility intervals of each estimated quantity had reached stable levels. I then confirmed that the Gelman–Rubin statistics were close to 1.0, as expected for chains that have converged to the target distribution (Gelman & Rubin 1992). The WinBUGS manual (Spiegelhalter et al. 2003) describes how these steps can be carried out, and further explanation is given by Gill (2002).

For the method of paired comparisons, probabilities of wins depend only on the differences between dominance abilities. The differences can be estimated by designating a focal individual, which is assigned a fixed dominance ability of 0, and estimating the dominance abilities of the other individuals relative to that benchmark. The individual chosen as the focal individual should be one of intermediate dominance ability and it is advisable to determine whether the inferences change substantially by designating others as the focal individual instead. Except for the focal individual, vague prior distributions for the d_i were set by assigning normal priors with $\mu = 0$ and $\sigma^2 = 10^3$. The large magnitude of σ^2 implies that plausible values of d_i are nearly equally likely. For some analyses in WinBUGS, it is necessary to impose further constraints on the values of d_i; for example, by requiring d_i values to lie between -15 and 15. Under the Bradley–Terry model, when one individual's dominance ability exceeds that of another animal by 15, then if they were to meet in 25 000 contests, the stronger individual would prevail in all of them with a probability greater than 0.99. Thus, this constraint allows very large differences in strength, sufficient to ensure victory in all observed encounters even for large sample sizes.

To conduct analyses not supported by WinBUGS, a supplemental program, available from the author, was written in Delphi 7.0 (Borland Software Corporation), for Windows (Microsoft Corporation) operating systems. This

program estimates the posterior probabilities of alternative rank orders and conducts posterior predictive checks, described below.

Example of a Bayesian Analysis of Ranks

To illustrate the Bayesian approach, data on five cockroaches (Bell & Gorton 1978) were reanalysed (Table 1; Appendix 1). This example is of interest because alternative methods suggest different ranks (de Vries 1998). Individual C was assigned a dominance ability of zero and the joint posterior distribution of the dominance abilities of the other four individuals was determined under the Bradley–Terry model. After discarding a burn-in of 500 steps, 10 000 samples were drawn from the posterior distribution.

Results are shown in the first columns of Table 3. Since the 95% credibility intervals for the d_i of individuals A and B do not span 0, one can be confident that these individuals have higher dominance abilities than individual C. Table 3 also shows the posterior probability for the most likely rankings, measured as the proportion of steps in the Markov chain that a particular ranking was encountered. The eight rankings shown constitute the 95% credibility set, which is the smallest set of rankings containing at least 95% of the posterior samples. An additional 15 orderings were encountered at lower frequencies.

This analysis assumes that the dominance ability of individual C is fixed at 0 and assigns normal prior probability distributions with variances of 10^3 to the other individuals. It is advisable to evaluate whether the results are sensitive to these specifications. By comparing estimates based on three vague prior distributions, all with means equal to 0, Table 3 shows that the main inferences depend very little on the choice of priors. Changing which individual is assigned the benchmark dominance index of 0 caused only small changes in estimates of

relative dominance abilities, or in the posterior probabilities of alternative rankings.

The adequacy of the statistical model can be assessed by determining whether one or more aspects of the data are surprising given the model. In posterior predictive checks, the data are compared to simulated data sets drawn from the posterior distribution (Gelman et al. 2004). As an example, the chi-square statistic was used to measure the goodness of fit between the data (Table 1) and the values expected in each cell of the matrix given the means from the posterior probability distributions. For the cockroach data, the chi-square test statistic is 7.2. To see whether this indicates a surprising lack of fit, d_i values were drawn at random from the posterior probability distribution and used to simulate a data set with the observed sample sizes (the n_{ij}) using the logit link. The chi-square goodness-of-fit statistic was calculated for each simulated data set, in the same manner as for the real data set; these values were sorted and stored. The Bayes P value is the proportion of chi-square values for simulated data sets that are smaller than the chi-square value for the observed data; thus, a large P value would indicate a surprisingly large discrepancy between the data and the model. In this case, the Bayes P value is 0.23, which provides no evidence of a surprising lack of fit.

The I&SI method was applied following the algorithm of de Vries (1998) with nTries, the number of times a new ordering is generated by swapping individuals, set to 10 000. This method identified a single rank ordering, B>A>C>E>D, as being most consistent with the data. The posterior probability of this ordering is approximately 0.05; thus, this Bayesian analysis indicates that there is only a 5% chance that this is the true ordering.

A supplemental program, available from the author, sums the posterior probability for alternative orderings with similar properties. For example, the posterior probability that the dominance structure is (A,B) > (C,D,E), which is the combined probability for all orderings in which A and B outrank C, D and E, is 0.98. The probability

Table 3. Sensitivity of the inferences to the specification of the prior probability distributions

	Prior probability distribution							
	Normal variance$=10^3$			Normal variance$=10^9$			t_2 distribution variance$=10^8$		
Means and 95% credibility intervals of the posterior distributions									
	2.5%	Mean	97.5%	2.5%	Mean	97.5%	2.5%	Mean	97.5%
d_A	0.27	1.15	2.08	0.27	1.15	2.08	0.30	1.16	2.09
d_B	0.05	0.91	1.84	0.05	0.91	1.84	0.06	0.92	1.84
d_C		0			0			0	
d_D	−1.78	−0.55	0.65	−1.78	−0.55	0.65	−1.73	−0.52	0.69
d_E	−2.00	−0.88	0.21	−2.00	−0.88	0.21	−2.00	−0.87	0.24
Posterior probabilities of the eight most likely rank orderings									
A>B>C>D>E	0.40			0.40			0.39		
A>B>C>E>D	0.18			0.18			0.17		
B>A>C>D>E	0.14			0.14			0.15		
A>B>D>C>E	0.11			0.11			0.11		
B>A>C>E>D	0.05			0.05			0.05		
B>A>D>C>E	0.04			0.04			0.04		
A>B>D>E>C	0.02			0.02			0.02		
A>B>E>C>D	0.02			0.02			0.02		

that one particular individual outranks another was estimated by the proportion of steps in the Markov chain in which the estimated dominance ability of the first individual was greater than that of its rival. For example, the probability that D outranks E is 0.73.

Example with a Covariate

In many cases, the primary goal is to test whether position in the dominance hierarchy is associated with some other attribute of the animals, rather than to determine the single most likely rank ordering (Leonard 1977; Tufto et al. 1998). Here, I illustrate Bayesian regression of dominance against a possible contributor to competitive ability, such as body size. Regression coefficients are estimated within the same model used to estimate the dominance abilities; thus, information on uncertainty in ranks is incorporated into inferences about the regression parameters. Bayesian regression analysis requires addressing many of the same issues that must be considered in standard regression analysis, including which explanatory variables should be included, whether relationships are linear or nonlinear, and detection of outliers (Congdon 2003). A full treatment is beyond the scope of this paper; instead, a simple example is illustrated.

Let X represent a covariate, such as age or body size, that may contribute to dominance ability. The regression component of the model expresses dominance ability as a function of the covariate:

$$d_i = \beta[X_i - \text{mean}(X)] + \epsilon$$

where X_i is the value of the covariate for the ith individual, $\text{mean}(X)$ is the mean of the observed covariate values, and ϵ is a normally distributed error with a mean of 0 and unknown variance σ^2. Centering the independent variable by subtracting the mean improves the efficiency of MCMC estimation (Congdon 2001). The regression parameters are assigned standard vague priors, which implies that plausible values are judged to be nearly equally likely before observing the data. It is not necessary to include an intercept or to designate a focal animal; instead, d is equal to 0 when the covariate value is equal to the mean of the sample. A nonzero intercept would add a constant to each d_i, which has no effect on the inferences, because the probabilities of wins are governed only by differences in dominance indexes.

For cases in which regression assumptions may not be warranted, the Spearman rank correlation coefficient between position in the hierarchy and the covariate is often reported. This statistic can be calculated taking into account uncertainty in ranks. To accomplish this, the Spearman correlation coefficient is calculated at each step in the Markov chain and the results are collected to determine the mean and 95% credibility interval. This is not a fully Bayesian approach, and has a more limited interpretation than the regression analyses, since it is an estimate only for the individuals in the sample, rather than for the larger population. This measurement can be used when it is not clear whether the covariate is a cause

or an effect of dominance or when the assumptions of regression analysis are not met.

As an example, data on bighorn sheep, *Ovis canadensis*, from Hass (1991) were analysed (Table 2). This data set has generally modest sample sizes and many missing dyads, typical of observational studies. The coefficient β was assigned a normal prior probability distribution with a mean of 0 and a variance of 1000. The standard deviation, σ, of variation around the regression line was assigned a prior distribution with a uniform density between 0 and 1000. Animals known to be 7 years old or older were assigned ages of 7. The specification of this model in WinBUGS is listed in Appendix 2. Convergence of the Markov chains to the posterior distribution was slower than that for the cockroach example. After discarding a burn-in of 10 000 steps, 50 000 values were sampled from the posterior distribution (Fig. 2).

The 95% credibility interval for β is confined to values well above 0 (2.0 to 4.06; mean = 3.05) (Fig. 2); therefore, there is strong evidence for an increase in dominance ability with age. The standard deviation of the error term, σ, has a mean of 2.60 (95% credibility interval 1.39 to 4.30). The ranks of the animals cannot be determined with confidence; indeed, the 95% credibility set included more than 45 000 possible orderings. No rank ordering was encountered in more than four of the 50 000 steps of the Markov chain; thus, none was shown to have a posterior probability greater than 0.0001. By summing posterior probabilities across similar rankings, I evaluated the probability of some simpler claims. For example, the posterior probability that all females that were at least 7 years of age outranked all younger females was 0.88.

The magnitude of the slope of dominance ability against age should be interpreted with caution, however, because the estimates of dominance abilities and of the slope, β, may be sensitive to the specification of the priors, especially for data sets in which one individual or subset
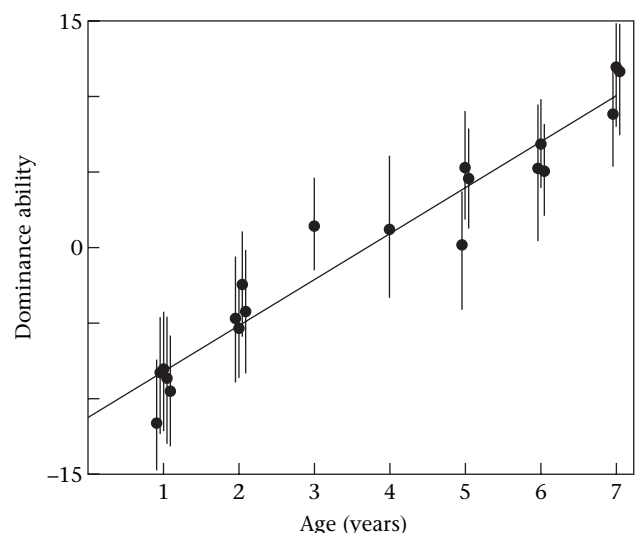

Figure 2. Regression of dominance ability against age for female bighorn ship, *Ovis canadensis*. Vertical bars represent 95% credibility intervals. The symbols are jittered along the age axis.

of individuals wins all observed encounters. For this example, substantial changes in the priors for β or σ had little effect on the posterior estimates of β. However, if the d_i values are constrained to lie within 12 units from that of the focal animal, instead of within 15 units, then the mean estimate of β declines from 3.05 to 2.55. Despite this change, the 95% credibility interval for β remains confined to positive values (1.71 to 3.32); thus, one can be confident that dominance increases with age. This is also shown by the Spearman correlation between dominance rank and age for this group of 20 females (mean = 0.88; 95% credibility interval: 0.80 to 0.93), taking into account uncertainty in ranks.

PERFORMANCE OF BAYESIAN AND I&SI METHODS

Simulation Methods

Use of nonparametric method to assign ranks is sometimes recommended when it is not clear that the assumptions of more specific models are met (de Vries 1998). To determine if this advice is warranted, I simulated data sets under a variety of rules, described below, governing relative dominance abilities and sample sizes, including cases in which the assumptions of the Bradley–Terry model are violated. I analysed each data set by the I&SI method and by Bayesian estimation of the Bradley–Terry model. The simulations and analyses were implemented in a program written in Delphi 7.0 (Borland Software Corporation). The dominance abilities of the animals and the number of encounters with other individuals were chosen by various rules. For each set of rules, 1000 data matrices were generated retaining information on each individual's true dominance ability. For each pair of individuals, the probability of a win by one of the individuals, π_{ij}, was determined using the logit link of the Bradley–Terry model (Fig. 1). For each encounter, a number between 0 and 1 was drawn at random and, if this number was smaller than π_{ij}, then the encounter was scored as a win for individual A_i, otherwise as a win for individual A_j. This was repeated until the entire data matrix was generated.

For Bayesian analysis, the prior probability distribution for all d_i values was normal with μ = 0 and $\sigma^2 = 10^2$. The joint posterior probability distribution of d_i was found by the Metropolis–Hastings algorithm, using normal proposal distributions with variances adjusted to achieve acceptance rates between 0.25 to 0.5 (Chib & Greenberg 1995). Because successive draws from the posterior sample were autocorrelated, the samples were thinned to retain only 1 in every 250 samples. For each data set, 1000 values were collected from the Markov chain after thinning and discarding a burn-in of 10. Preliminary analyses confirmed that this analysis yields the same posterior estimates as the WinBUGS package. Gelman–Rubin convergence diagnostics were less than 1.004; thus, there was no evidence of lack of convergence to the posterior distribution. Because the number of steps in each Markov chain was modest, the simulations produced conservative estimates of the

accuracy of Bayesian inferences. The I&SI method was carried out following the algorithm of de Vries (1998) with nTries, the number of times a new ordering is generated by swapping individuals, set to 10 000. Preliminary trials showed that the performance of the I&SI method was not improved substantially by using larger values of nTries.

The following results were collected: whether or not the method detected the correct rank order unambiguously (that is, without tying other rankings in the degree of support), and the Spearman correlation between the estimated ranks and the true ranks. If the Bayesian or I&SI method produced several rank orderings that tied for best, then the average Spearman correlation coefficient for the tied orderings was calculated.

I varied the following attributes across simulations.

(1) The number of individuals was either 5, 10, or 15.

(2) In each simulation, each individual's dominance ability was drawn randomly from a normal distribution with a mean of 0. The standard deviation of this distribution was varied across simulations to produce groups with modest differences in competitive ability, achieved by setting the SD to 0.5, or with larger differences in competitive ability, achieved by setting the SD to 5.0. When the SD was equal to 0.5, the average probability of a win over an individual one rank lower was 0.54 to 0.60. When the SD was equal to 5.0, the average probability of a win over an individual one rank lower was 0.62 to 0.74.

(3) The sample sizes of encounters per pair were either even, uneven, or weighted towards higher-ranked individuals ('top-heavy'), but with the same average number of observations per pair. For even sample sizes, n_{ij} was set to 10 for each pair. For uneven sample sizes, the n_{ij} for each pair was drawn randomly from the integers ranging from 0 to 20, to yield an average of 10. For top-heavy sample sizes, the n_{ij} for each pair was adjusted so that there were more observations for the highest-ranking individuals, a common result of observational studies. Specifically, the sample size for each pair was weighted according to the product of their dominance indexes raised to an exponent, the value of which was set so that 5–6% of pairs had $n_{ij} = 0$. These weights were multiplied by a common factor so that the number of observations, averaged across all pairs, was as close as possible to 10.

I conducted a second set of simulations in which the assumptions of the Bradley–Terry model were violated to determine whether nonparametric methods are more reliable in these cases. Each result is based on 1000 simulations of five individuals with dominance abilities drawn at random from a normal distribution with a mean of 0 and a variance of 0.5. Sample sizes per pair averaged 10, but were uneven (see above). Four types of assumption violations were implemented.

(1) Variable dominance abilities

Animals' competitive abilities in natural systems may vary from encounter to encounter. This was modelled by assigning each individual a mean, μ_i, drawn from a normal distribution with a mean of 0 and a variance of 0.5.

However, instead of always using the same value of d_i for a given individual, for each encounter, d_i was drawn from a normal distribution with mean $= \mu_i$ and a variance of either 1, 2, 4, or 8. Notice that these rules cause the variation in dominance abilities through time for the same individual to be large compared with the variation between individuals.

(2) Pair-specific effects

The Bradley–Terry model assumes that each animal is characterized by the same dominance ability regardless of its opponent. To vary this assumption, for each pair of individuals, the difference in dominance ability was adjusted by adding an extra term specific to that pair. This allows nontransitive dominance relationships. The extra term was drawn randomly from a normal distribution with a mean of 0 and a standard deviation of 0.5, 1.0, or 1.5.

(3) Autocorrelated wins and losses

To simulate autocorrelated outcomes for successive encounters between a pair of individuals, the outcome of the first encounter was generated at random, based on the animals' dominance indexes and the link function, then repeated for two, four, or eight consecutive steps. This procedure was repeated up to the limit of the number of observations for that pair. For example, when outcomes were repeated for four steps, the results of 10 encounters between two equally matched animals A and B might be (A, A, A, A, B, B, B, B, A, A), where the letter represents the winner.

(4) Alternative link functions

To vary the assumption that the probability of a win has a logistic relationship to differences in dominance abilities, data were simulated by using a probit or a cumulative Cauchy link function, each centred at 0 (Fig. 1). The results were analysed using the Bradley–Terry model, to simulate ignorance by the investigator of the true function.

Simulation Results

Bayesian estimation outperformed the I&SI method for all sets of simulations by both criteria: the percentage of replicates in which the correct ranks were determined unambiguously was higher, and the average correlation between the estimated and true ranks was higher (Tables 4, 5).

For most simulated data sets, no method was likely to identify unambiguously the correct dominance order (Table 4). The success rate was higher if the number of individuals was small, if differences in dominance abilities were comparatively large, and if sample sizes per pair were even. However, many published data sets either have a larger number of individuals, smaller differences in dominance ability between at least some pairs, and uneven sample sizes. In these cases, simulations showed that the

Table 4. Performance of the Bayesian model and of the I&SI method in detecting the true structure of dominance hierarchies in simulated data sets

Number of animals	Variance in strengths	Sample sizes	% Determined*		Correlation†	
			Bayesian	I&SI	Bayesian	I&SI
5	0.5	Even	27.4	10.1	0.827	0.773
		Uneven	25.9	10.9	0.818	0.748
		Top-heavy	24.9	3.1	0.786	0.673
10	0.5	Even	2.1	0	0.913	0.868
		Uneven	1.2	0	0.902	0.838
		Top-heavy	0.7	0	0.877	0.758
15	0.5	Even	0	0	0.931	0.896
		Uneven	0	0	0.927	0.878
		Top-heavy	0	0	0.879	0.793
5	5.0	Even	60.7	43.1	0.949	0.935
		Uneven	54.2	35.5	0.933	0.904
		Top-heavy	43.3	15.1	0.899	0.847
10	5.0	Even	19.0	2.9	0.974	0.958
		Uneven	15.9	2.1	0.970	0.947
		Top-heavy	5.5	0.1	0.945	0.892
15	5.0	Even	3.0	0	0.983	0.970
		Uneven	2.9	0	0.982	0.963
		Top-heavy	0.3	0	0.965	0.930

*Percentage of replicates in which the true rank order is unambiguously determined.
†Rank correlation between the true and the estimated ranks.

Table 5. Performance of Bayesian analysis and the I&SI method when the assumptions of the Bayesian model are violated

	% Determined*		Correlation†	
	Bayesian	I&SI	Bayesian	I&SI
All assumptions met				
	25.9	10.9	0.818	0.748
Dominance abilities vary across encounters				
Variance in dominance abilities				
1.0	18.6	6.8	0.737	0.633
2.0	15.7	5.7	0.705	0.629
4.0	9.7	3.1	0.618	0.531
8.0	7.4	2.8	0.545	0.472
Pair-specific effects				
Variance in added term				
0.5	14.8	5.1	0.718	0.643
1.0	10.0	4.8	0.604	0.544
1.5	7.4	3.8	0.581	0.521
Autocorrelated outcomes				
Number of repeated results				
2	18.4	7.6	0.757	0.663
4	10.7	5.0	0.631	0.569
8	6.8	3.6	0.540	0.492
Alternative link functions				
Probit	41.3	18.5	0.880	0.827
Cauchy	28.9	12.3	0.832	0.775

*Percentage of replicates in which the true rank order is unambiguously determined.
†Rank correlation between the true and the estimated order.

probability of detecting the true rank order unambiguously was often less than 5% (Tables 4, 5).

Violations of the model assumptions usually reduced the probability that the correct rank order was detected by Bayesian analysis (Table 5). However, the performance of the I&SI method also suffered when the same assumptions were violated, and in all cases investigated, the Bayesian method outperformed the I&SI method. When the data were generated using a probit or cumulative Cauchy link function (Fig. 1), but analysed assuming a logit link, the ability to detect the true rankings and correlation coefficients was higher than when the data were generated using a logit link (Table 5). This is because small differences in dominance indexes result in a higher proportion of encounters won by the stronger individual under the probit and Cauchy models than under the Bradley–Terry model (Fig. 1).

DISCUSSION

Bayesian analysis of dominance hierarchies offers significant advantages over methods that identify ranks by nonparametric criteria, such as the I&SI method (de Vries 1998; de Vries & Appleby 2000). Simulations showed that the inferences of Bayesian versions of the Bradley–Terry model are more accurate than those of the I&SI method under a variety of circumstances, even when the assumptions of the Bradley–Terry model are violated (Tables 4, 5). Bayesian analyses are also more informative, since they produce measures of certainty, or uncertainty, about the inferences. For example, in addition to determining the most likely ordering, the models described here yield the posterior probability for that ordering (e.g. Table 3). Furthermore, information on uncertainty of ranks is retained in further steps of the analysis, such as regressions of dominance ability against other attributes of the animals. Thus, questions about the causes and consequences of positions within a hierarchy may be addressed rigorously even when there is substantial uncertainty as to the ranks themselves.

The Bayesian approach also solves a major limitation that otherwise prevents application of the Bradley–Terry model to many data sets. When maximum likelihood methods are used, estimates of dominance abilities do not converge for any data set in which an individual or group of individuals has prevailed in all observed encounters (Boyd & Silk 1983). In the Bayesian versions, incorporation of prior probability distributions allows estimation of the Bradley–Terry model for any data set (Leonard 1977).

A potential disadvantage of Bayesian analysis is the need to specify prior distributions when little information is available for guidance. Bayesian inferences are influenced both by the priors and the data (Gill 2002); however, the impact of the priors declines as data accumulate and so key inferences are often insensitive to the choice of priors (e.g. Table 3). Yet with very small sample sizes, the priors can have a larger effect on the inferences than in the examples presented here. Post hoc analysis of the sensitivity of the estimates to the priors is recommended and the priors should be no more specific than is warranted by biological knowledge.

A second possible disadvantage of Bayesian analysis is the greater time needed for computation. The MCMC techniques used by the software to estimate the joint posterior distribution of the unknown parameters are powerful, but care is needed to ensure that the Markov chains converge to the target distribution. In practice, this means that the properties of one or more preliminary runs must be examined before the final values are generated.

Although Bayesian analysis of dominance hierarchies appears to make more assumptions than do the various nonparametric techniques, the simulations summarized in Table 5 show that the I&SI method is at least as sensitive as the Bayesian method to these assumptions. These include the assumption that dominance abilities are constant, that wins and losses in successive observations are not correlated, and that the probabilities of wins are determined only by differences in dominance indexes and not by additional terms specific to each pair. To the extent that these assumptions are violated, Bayesian analyses are less able to determine the true ranks, but this is also true for nonparametric methods. For all of the simulations conducted for this study, the results of the Bayesian analysis were on average closer to the true values (Table 5). When sample sizes are uneven or correlated with dominance ability, this is reflected by variation in the breadths of the 95% credibility intervals, but it does not reduce the accuracy of Bayesian inferences relative to nonparametric techniques (Table 4).

Furthermore, model checking often allows one to detect which assumptions are violated, allowing the underlying model to be modified (Gelman et al. 2004). Posterior predictive checks reveal a surprising lack of fit of the model to the data when there are strong intransitive relationships between individuals, or when observed values are overdispersed compared with the expectations of the model. Alternatively, the fit of alternative models allowing intransitive relationships or different types of transitivity can be compared (Tufto et al. 1998).

Bayesian analysis of dominance hierarchies is not limited to the simple models discussed in this paper. When additional information is present, such as the order in which wins and losses occur, Bayesian methods may be used to infer the process determining an animal's position in the hierarchy and how dominance ability depends on previous wins and losses.

References

Bell, W. J. & Gorton, R. E., Jr. 1978. Informational analysis of agonistic behaviour and dominance hierarchy formation in a cockroach, *Nauphoeta cinerea*. *Behaviour*, **67**, 217–235.

Boyd, R. & Silk, H. B. 1983. A method for assigning cardinal dominance ranks. *Animal Behaviour*, **31**, 45–58.

Bradley, R. A. & Terry, M. E. 1952. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324–345.

Chib, S. & Greenberg, E. 1995. Understanding the Metropolis–Hastings algorithm. *American Statistician*, **49**, 327–335.

Congdon, P. 2001. *Bayesian Statistical Modelling*. Chichester: J. Wiley.

Congdon, P. 2003. *Applied Bayesian Modelling*. Chichester: J. Wiley.

David, H. A. 1988. *The Method of Paired Comparisons*. London: Charles Griffen.

Davidson, R. R. & Solomon, D. L. 1973. A Bayesian approach to paired comparison experimentation. *Biometrika*, **60**, 477–487.

Drews, C. 1992. The concept and definition of dominance in animal behavior. *Behaviour*, **125**, 283–313.

Gelman, A. & Rubin, D. B. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. 2004. *Bayesian Data Analysis*. Boca Raton, Florida: Chapman & Hall.

Gill, J. 2002. *Bayesian Methods: a Social and Behavioral Sciences Approach*. Boca Raton, Florida: Chapman & Hall.

Hass, C. C. 1991. Social status in female bighorn sheep (*Ovis canadensis*): expression, development and reproductive correlates. *Journal of Zoology*, **225**, 509–523.

Hilborn, R. & Mangel, M. 1997. *The Ecological Detective: Confronting Models with Data*. Princeton, New Jersey: Princeton University Press.

Holder, M. T. & Lewis, P. O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, **4**, 275–284.

Holsinger, K. E., Lewis, P. O. & Dey, D. K. 2002. A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, **11**, 1157–1164.

Leonard, T. 1977. An alternative Bayesian approach to the Bradley–Terry model for paired comparisons. *Biometrics*, **33**, 121–132.

Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. 2003. *WinBUGS User Manual*. Version 1.4. Cambridge: MRC Biostatistics Unit.

Tufto, J., Solberg, E. & Ringsby, T. 1998. Statistical models of transitive and intransitive dominance structures. *Animal Behaviour*, **55**, 1489–1498.

de Vries, H. 1998. Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. *Animal Behaviour*, **55**, 827–843.

de Vries, H. & Appleby, M. C. 2000. Finding an appropriate order for a hierarchy: a comparison of the I&SI and the BBS methods. *Animal Behaviour*, **59**, 239–245.

Appendix 1

WinBUGS specification of the Bradley–Terry model for the data in Table 1

```
model {

    # For each of the 9 dyads
    for (i in 1:9) { n[i] <- win1[i] + win2[i]

      # the number of wins by individual 1 is a binomial
      # random variable
      win1[i] ~ dbin (p[i],n[i])
      # the logit of p is equal to the difference between
      # the dominance indexes
      logit(p[i]) <- d[ind1[i]] − d[ind2[i]] }


      # d is given a fixed value for individual 3; normal
      # priors for the others
```

```
    # normal distributions are parameterized by the mean
    # and the reciprocal of the variance
    d[3] <- 0
    for (i in 1:2) { d[i] ~ dnorm(0,0.001) }
    for (i in 4:5) { d[i] ~ dnorm(0,0.001) }


}
# Initial values for the Markov chain
list(d = c(0,0,NA,0,0))
Data
# ind1 and ind2 are indexes of the two animals;
# win1 and win2 are numbers of wins
```

ind1[ ]	ind2[ ]	win1[ ]	win2[ ]
1	2	9	10
1	3	12	2
1	4	6	3
1	5	27	2
2	3	9	5
2	4	12	3
2	5	12	3
3	5	2	0
4	5	2	4
END			

Appendix 2

WinBUGS specification of the bighorn sheep regression

```
model {

    # Interactions were observed for 96 dyads
    for (i in 1:96) {

      n[i] <- win1[i] + win2[i]
      win1[i] ~ dbin(p[i],n[i])
      logit(p[i]) <- d[ind1[i]] − d[ind2[i]]

    }
    for (i in 1:20){

      d[i] ~ dnorm(mu[i],tau) I(−15,15)
      # The regression component
      mu[i] <-beta*(age[i]−mean(age[ ]))

    }
    # Specification of priors
    beta ~ dnorm(0,0.001)
    tau <- pow(sigma,−2)
    sigma ~ dunif(0,1000)

}
Initial values for the Markov chain
list(beta = 0, sigma = 1, d = c(0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0))
Data on the covariate
list(age = c(7,7,7,5,6,6,4,6,3,5,2,5,2,2,2,1,1,1,1,1))
```

For brevity, only the first two nonzero dyads from Table 2 are shown. No entry is made for any pair for which no encounters were observed.

ind1[ ]	ind2[ ]	win1[ ]	win2[ ]
1	3	4	0
1	5	7	0
.			
.			
END			

```
# Alternative code to find the Spearman rank correlation
# coefficient between dominance and covariate age. The
# following should be substituted for the loop in the
# code above containing the regression component.
# d is given a fixed value for individual 7
d[7] <- 0
for (i in 1:6){
```

```
   d[i] ~ dnorm(0,tau) I(−15,15)

}
for (i in 8:20){
   d[i] ~ dnorm(0,tau) I(−15,15)
}
# calculate the Spearman rank correlation coefficient
for (i in 1:20){
   dsqr[i] <- pow((rank(age[ ],i) − rank(d[ ],i)),2)
}
sumdsqr <- sum(dsqr[ ])
Spearman <- 1 − ((6∗sumdsqr)/(pow(20,3) − 20))
```