



Testing association patterns: issues arising and extensions

HAL WHITEHEAD, LARS BEJDER & C. ANDREA OTTENSMEYER

Department of Biology, Dalhousie University, Halifax

(Received 4 May 2004; initial acceptance 30 August 2004;
final acceptance 18 November 2004; published online ■ ■ ■; MS. number: AF-26)

A reasonable necessary condition for a population to be socially structured is that individuals associate in a nonrandom fashion. Thus, testing for preferred or avoided companions is a fundamental step in analyses of social structure (Whitehead & Dufault 1999). Cluster analyses, sociograms and their ilk (e.g. Morgan et al. 1976) all assume nonrandom associations, and if associations are random, none of these mean anything.

Unfortunately testing nonrandom association on real data is not simple. A suitable test statistic is the standard deviation of the association indexes, which will be higher than expected if individuals have preferred or avoided associates. But what is 'expected' in the case of no preference or avoidance? The distributions of the standard deviation of association indexes, or any other suitable test statistic, are not analytically tractable under the null hypothesis. A solution is to use permutation tests in which the association data are randomized many times subject to certain constraints, each time calculating the test statistic (Manly 1997). The mean of these randomized test statistics can be considered its 'expected value', and a *P* value is then calculated as the proportion of times that the permuted statistics are more extreme than the real test statistic. A number of analyses of animal social structure have taken this approach (e.g. Whitehead et al. 1982; Smolker et al. 1992; Slooten et al. 1993; Pepper et al. 1999). However, it was not easy to design an efficient computational routine that randomly permutes the records of which individuals were found in which groups in such a way that the number of individuals in each group and the number of groups containing each individual are both held constant. There are conceptual difficulties with such tests (Manly 1997).

However, using a routine developed by Manly (1995) for an ecological problem, Bejder et al. (1998) showed how an individual by group 1:0 matrix could be permuted simply while keeping constant both the number of animals in each group and the number of groups in which each animal was identified. In their procedure (referred to as the MBFB method), at each step, two individuals and two groups are chosen so that each individual is identified in just one of the groups, and each group contains just one of

the individuals. Then the four group-individual assignments are switched (the individual in group A is now in group B and vice versa; Table 1). We will call this a 'flip'. The flipping procedure preserves the totals for each group and individual. Permutations produced with flips are not independent, but Manly (1995) showed that this does not matter as long as sufficient flips are carried out (typically many greater than the 1000 or so that are usual for permutation tests). In the MBFB method, after each flip in the group-individual matrix, the test statistic is calculated, and so a distribution of 'random' test statistics is produced from which the *P* value is calculated.

If animals are recruited to the population, die, or migrate to or from the study area during the study, then the MBFB method could suggest significant preference/avoidance just because some pairs of animals shared time together in the study area and others did not. To remove such demographic effects, Whitehead (1999) suggested that the flips only be carried out within temporal 'sampling periods' within which demographic events are unlikely (Table 2). Another possible modification (Whitehead 1999) is to make flips within 1:0 symmetric association matrices (symmetric in the sense that the association between individuals A and B is the same as that between B and A) for each sampling period (Table 3), thus permitting tests for preferred/avoided companionship between sampling periods when groups are not defined but associations are defined (e.g. from nearest-neighbour distances; temporal synchrony in behaviour; or dyadic interactions such as grooming behaviour). We call the original MBFB method total-matrix permutation, and Whitehead's (1999) two modifications within-period permutation and between-period permutation, respectively.

One useful attribute of the MBFB method and its modifications is that *P* values can also be calculated for each dyad, so that it is possible to test whether the members of that particular dyad associate preferentially, or avoid one another, against the null hypothesis that there is no particularly strong or weak association. This is done by calculating the proportion of random dyadic association indexes (i.e. after each flip) that are greater than the real association index for that dyad.

The MBFB method has been used in 14 peer-reviewed published papers on species ranging from Spix's disc winged bats, *Thyroptera tricolor*, to chimpanzees, *Pan troglodytes*, to sperm whales, *Physeter macrocephalus*. Of

Table 1. One flip in the process of permuting a group–individual matrix by the total-matrix permutation method (from Whitehead 1999)

Individuals				Individuals		
Group	A B C D E F G	⇒	Group	A B C D E F G	Group	A B C D E F G
a	1 1 0 0 0 0 0	⇒	a	1 1 0 0 0 0 0	a	1 1 0 0 0 0 0
b	1 1 1 1 0 0 0		b	1 1 1 1 0 0 0	b	1 1 1 1 0 0 0
c	1 1 0 0 0 0 0		c	1 1 0 0 0 0 0	c	1 1 0 0 0 0 0
d	1 1 1 1 0 0 0		d	1 1 1 1 0 0 0	d	1 1 1 1 0 0 0
e	0 0 1 0 1 1 1		e	0 0 1 0 1 1 1	e	0 0 1 0 1 1 1
f	0 0 1 0 0 1 1		f	0 0 1 0 0 1 1	f	0 0 1 0 0 1 1
g	0 0 0 0 1 1 1		g	0 0 0 0 1 1 1	g	0 0 0 0 1 1 1
h	0 0 0 0 0 1 1		h	0 0 0 0 0 1 1	h	0 0 0 0 0 1 1
i	0 1 0 0 0 0 1		i	0 1 0 0 0 0 1	i	0 1 0 0 0 0 1
j	0 1 0 0 0 0 1		j	0 1 0 0 0 0 1	j	0 1 0 0 0 0 1
k	0 1 0 0 0 1 0		k	1 1 0 0 0 0 0	k	1 1 0 0 0 0 0
l	1 0 1 0 0 0 0		l	1 0 1 0 0 0 0	l	1 0 1 0 0 0 0
m	0 0 1 0 1 0 0		m	0 0 1 0 1 0 0	m	0 0 1 0 1 0 0
n	1 0 0 0 1 0 0		n	1 0 0 0 1 0 0	n	1 0 0 0 1 0 0
o	1 0 0 0 1 0 0		o	1 0 0 0 1 0 0	o	1 0 0 0 1 0 0
p	0 0 1 1 0 0 0		p	0 0 1 1 0 0 0	p	0 0 1 1 0 0 0
q	1 0 0 1 0 0 0		q	1 0 0 1 0 0 0	q	1 0 0 1 0 0 0
r	1 0 1 1 0 0 0		r	0 0 1 1 0 1 0	r	0 0 1 1 0 1 0
s	1 0 1 1 1 0 0		s	1 0 1 1 1 0 0	s	1 0 1 1 1 0 0
t	0 0 1 0 1 0 0		t	0 0 1 0 1 0 0	t	0 0 1 0 1 0 0

The matrix on the left, showing which groups contained which individuals, was modified by randomly choosing two individuals and two groups (with each individual in only one of the groups and each group containing only one of the individuals) and switching assignments (bold) to preserve row and column totals.

Table 2. One flip in the process of permuting a group–individual matrix by the within-period permutation method in which demographic effects are controlled (from Whitehead 1999)

Individuals				Individuals		
Group	A B C D E F G	⇒	Group	A B C D E F G	Group	A B C D E F G
a	1 1 0 0 0 0 0	⇒	a	1 1 0 0 0 0 0	a	1 1 0 0 0 0 0
b	1 1 1 1 0 0 0		b	1 1 1 1 0 0 0	b	1 1 1 1 0 0 0
c	1 1 0 0 0 0 0		c	1 1 0 0 0 0 0	c	1 1 0 0 0 0 0
d	1 1 1 1 0 0 0		d	1 1 1 1 0 0 0	d	1 1 1 1 0 0 0
e	0 0 1 0 1 1 1		e	0 0 1 0 1 1 1	e	0 0 1 0 1 1 1
f	0 0 1 0 0 1 1		f	0 0 1 0 0 1 1	f	0 0 1 0 0 1 1
g	0 0 0 0 1 1 1		g	0 0 0 0 1 1 1	g	0 0 0 0 1 1 1
h	0 0 0 0 0 1 1		h	0 0 0 0 0 1 1	h	0 0 0 0 0 1 1
i	0 1 0 0 0 0 1		i	0 1 0 0 0 0 1	i	0 1 0 0 0 0 1
j	0 1 0 0 0 0 1		j	0 1 0 0 0 0 1	j	0 1 0 0 0 0 1
k	0 1 0 0 0 1 0		k	0 1 0 0 0 1 0	k	0 1 0 0 0 1 0
l	1 0 1 0 0 0 0		l	1 0 1 0 0 0 0	l	1 0 1 0 0 0 0
m	0 0 1 0 1 0 0		m	0 0 1 0 1 0 0	m	0 0 1 0 1 0 0
n	1 0 0 0 1 0 0		n	1 0 0 0 1 0 0	n	1 0 0 0 1 0 0
o	1 0 0 0 1 0 0		o	1 0 0 0 1 0 0	o	1 0 0 0 1 0 0
p	0 0 1 1 0 0 0		p	0 0 1 1 0 0 0	p	0 0 1 1 0 0 0
q	1 0 0 1 0 0 0		q	0 0 0 1 1 0 0	q	0 0 0 1 1 0 0
r	1 0 1 1 0 0 0		r	1 0 1 1 0 0 0	r	1 0 1 1 0 0 0
s	1 0 1 1 1 0 0		s	1 0 1 1 1 0 0	s	1 0 1 1 1 0 0
t	0 0 1 0 1 0 0		t	1 0 1 0 0 0 0	t	1 0 1 0 0 0 0

The data were collected in five sampling periods, separated by line spaces. At this step, the fifth sampling period was randomly chosen, and within it, the matrix was modified by randomly choosing two individuals and two groups and switching assignments (bold), to preserve row and column totals within the sampling periods.

Table 3. One flip in the process of permuting an association matrix by the between-period permutation method (from Whitehead 1999)

Individual				Individual	
A B C D E F G H	⇒	A B C D E F G H	A B C D E F G H		
A	— 1 0 0 0 1 0 1	⇒	A	— 1 0 1 0 1 0 0	
B	1 — 1 1 0 0 1 1		B	1 — 1 1 0 0 1 1	
C	0 1 — 1 1 1 1 0		C	0 1 — 0 1 1 1 1	
D	0 1 1 — 0 1 0 0		D	1 1 0 — 0 1 0 0	
E	0 0 1 0 — 0 1 1		E	0 0 1 0 — 0 1 1	
F	1 0 1 1 0 — 0 0		F	1 0 1 1 0 — 0 0	
G	0 1 1 0 1 0 — 1		G	0 1 1 0 1 0 — 1	
H	1 1 0 0 1 0 1 —		H	0 1 1 0 1 0 1 —	

A sampling period is randomly chosen, and the associations within the sampling period, shown in the symmetric 1:0 matrix on the left, are modified by randomly choosing two pairs of individuals (D and H; A and C) so that each individual is associated with only one member of the other pair. The associations between these pairs (bold) are switched, preserving row (and column) totals.

these implementations of the MBFB method, 13 have used the MATLAB program package SOCPROG, written by one of us (H.W.). Twelve, eight and two published papers have used the total-matrix permutation, the within-period permutation and the between-period permutation methods, respectively.

Since writing the original papers on the MBFB method (Bejder et al. 1998) and its modifications (Whitehead 1999), and having reviewed studies that use the method, we have become aware of some important conceptual and computational issues, and have developed two extensions to the method. These are considered in this paper, and, in most cases, are incorporated into the latest version of SOCPROG (<http://is.dal.ca/~hwhitehe/social.htm>).

Conceptual Issues

Test statistics

Bejder et al. (1998) used the median of half-weight association (Cairns & Schwager 1987; Ginsberg & Young 1992) indexes as their test statistic (calculated for the real data, and after each flip). This could be appropriate if the alternative to the null hypothesis is that there are more dyads with strong relationships than would be expected when animals associated randomly. However, medians are computationally expensive to calculate and we may also be interested in whether there are dyads that actively avoid one another. In 'sparse' data sets, with many individuals and rather few groups, the median will often be zero, making the test redundant. Other test statistics may be more powerful, and sometimes the use of two or more statistics may be warranted. For instance, when using the within-period variant of the MBFB method (Whitehead 1999), it may be possible to use two different test statistics to examine randomness of association within and between sampling periods, respectively.

We examined the performance of five possible test statistics for use in testing nonrandom associations within and between sampling periods: (1) the mean of all pairwise association indexes (AIs); (2) the standard

deviation of all AIs; (3) the proportion of nonzero AIs; (4) the mean of nonzero AIs; and (5) the standard deviation of nonzero AIs. We used the 'simple-ratio' association index (AI) throughout (Cairns & Schwager 1987; Ginsberg & Young 1992).

Data sets were generated in MATLAB (the Mathworks, Natick, Massachusetts, U.S.A.) to simulate identification records for a population of 150 individuals, observed in groups over 15 sampling periods. In each sampling period, the number of groups observed was Poisson distributed with a mean of 7, and all individuals in each observed group were identified. Data sets with these characteristics were constructed to simulate three social scenarios: (1) random associations; (2) short-term preferences; and (3) long-term companionships. In the first scenario, each observed group was equally likely to contain 10, 20 or 30 individuals, with individuals randomly selected for each group. To simulate short-term preferences, individuals were assigned to 'parties' of 10 individuals for each sampling period; each group was equally likely to be composed of one, two or three parties, with parties randomly selected for each group. To simulate long-term companionships, individuals were assigned to 'units' of 10 individuals whose membership persisted across all sampling periods (the whole study); each group was equally likely to be composed of one, two or three units, with units randomly selected for each group. Ten simulated data sets were generated under each social scenario. Both the within-period and between-period permutation methods were used on each simulated data set, and P values calculated for each of the five test statistics (1000 permutations). Useful test statistics were those that had consistently low P values when the tested attribute was present in the data set, but few false positives (indicated by $P < 0.05$) when it was not.

The best indicators of nonrandom relationships within periods appear to be an unexpectedly low proportion of nonzero AIs (10/10 P values < 0.02 when testing short-term preference data sets, no false positive findings when testing the random associations data sets) and an unexpectedly low mean of all AIs (10/10 P values < 0.01 when testing short-term preference data sets, two false positive findings (0.039, 0.049) when testing random associations data sets). If there are preferred or avoided companions within periods, then the real statistics should be low compared with the 'random' statistics generated by permuting data sets within periods. This is because in the real data there will be proportionally more pairs of individuals that are repeatedly grouped (redundantly from the perspective of calculating AIs) and thus, because the number of animals in each group is constrained, proportionally fewer pairs that are grouped, so decreasing the proportion of nonzero AIs, and the mean association index. Our simulations suggested that the mean of nonzero AIs may not be reliable as an indicator of within-period preferences when between-period preferences are also present (10/10 P values < 0.02 when testing short-term preference data sets, but highly variable P values when testing long-term companionship data sets).

A standard deviation of nonzero AIs that is higher than the 'random' nonzero standard deviations of data sets

permuted between periods may be the best indicator of companionships that persist across sampling periods (10/10 P values < 0.01 when testing long-term companionship data sets, no false positives for random or short-term preference data sets). The standard deviation of all AIs was also successful in indicating when these longer-term companionships existed (10/10 P values < 0.01), but was more prone to false positive readings when calculated for the random and short-term preference data sets. The possibility of false positive findings uncovered for some of the test statistics underscores the importance of repeating the permutation tests a few times on the same data set to check for stability.

We have recently realized that a low mean of the AIs, produced by short-term preferences, will also tend to lower the standard deviation of the AIs, and thus mask the presence of long-term preferences. Using the coefficient of variation of the AIs (standard deviation divided by mean) as a test statistic for long-term preferences, in place of their standard deviation, should compensate for this effect, but we have yet to investigate this.

Dyadic P values

In a few cases in the published literature, as well as in unpublished studies that we have reviewed, the dyadic P values from the MBFB method have been considered to be measures of the strength of an association between pairs of individuals and used as the data for further analyses (e.g. Lusseau et al. 2003 constructed a sociogram using dyadic P values). This is conceptually invalid because the size of a P value depends not only on the strength of the association between two individuals but also on the amount of available data (number of groups and/or sampling periods examined), and, in the case of the MBFB method and its modifications, the structure of the data: there may be few possible flips that use the data for some dyads, and many possibilities for others, affecting potential ranges and powers of dyadic P values. For this reason, coupled with the computational challenges of obtaining precise dyadic P values (see below), dyadic P values should not be used as a measure of the strength of the relationship between two animals. This is the purpose of the association index itself, and the dyadic P values are only guides as to the weight that should be placed on particularly high or low indexes (Johnson 1999). Many standard association indexes are affected by sample size, so that a pair of individuals seen often will have a higher association index than those seen rarely, even though each pair might have the same probability of being together given that they are both observed. To remove such effects, one can use an association index that does not have this property (e.g. the number of sampling periods during which a dyad was associated divided by the total number of sampling periods both individuals were identified; Christal & Whitehead 2001), or the ratio of observed to expected values of the association index (Pepper et al. 1999), with expected values calculated by randomization, perhaps using the MBFB method. The latter approach also standardizes for overall differences in sociality among individuals.

Computational Issues

Number of flips per permutation

In the original MBFB method, association indexes and test statistics are calculated after each nonindependent flip. Because a large part of the time taken for each flip is in the calculation of association indexes and test statistics, we examined the possibility of only making these calculations every n flips: the routine makes n flips, then calculates the association indexes and test statistics, giving fewer but less autocorrelated random test statistics.

For a variety of randomly produced data sets, and 1, 10, 100 or 1000 flips per permutation, we calculated the overall (using the SD of simple-ratio association indexes) and dyadic P values (p_j , $j = 1, \dots, r$) for $r = 5$ or $r = 10$ replicate runs ($r = 5$ runs for more time-intensive analyses) over a set amount of time (t , in seconds). These P values were compared with 'baseline' P values from running the routine with 100 flips per permutation, over $100 \times t$ s (P_B). The root mean square errors (RMSE) were calculated as:

$$\text{RMSE} = \sqrt{\left(\sum (p_j - P_B)^2 / r\right)}$$

RMSEs were calculated for both the overall test statistic, and the dyadic P values (taking the mean over all dyads). The results are presented in Fig. 1.

These results show that for the SOCPROG routines implemented on MATLAB, using anywhere between 10 and 1000 flips per permutation is much more efficient than calculating tests statistics after each permutation (Fig. 1). The optimal number seems to vary between data sets and analytical techniques, and will also vary with the computational routines used, but 100 seems to be a good general compromise for the SOCPROG routines. Using 10–1000 flips per permutation is generally very much more efficient (in terms of P value precision per unit of processing time) than the original procedure with one flip per permutation.

Calculation of dyadic P values

It is clear from Fig. 1 that dyadic P values are often, but not always, considerably less precise than overall P values. This results from each flip only affecting two individuals (or four in the case of flipping symmetric association matrices). For instance, with 32 individuals in a population, on average only one in every 16 flips (or one in eight flips with symmetric association matrices) will affect a particular dyadic P value. Users should be aware of this, and only use dyadic P values when the number of permutations is sufficient to stabilize dyadic P values.

Extensions

Between classes

The MBFB method and its modifications examine the possibility of preferred/avoided associations within symmetric association matrices; in other words, among all individuals in a defined population. Individuals can often be allocated to classes, by age, gender, mitochondrial

haplotype or some other attribute. Schnell et al. (1985) showed how to use the Mantel test to investigate whether associations between individuals of different classes are different from those between members of the same class. This question can also be addressed when animals are not individually identified but can be allocated to classes (e.g. Underwood 1981). A different, but related, issue introduced by Hemelrijk (1990) is a test for associations between two classes of individuals. For instance, one might ask: 'do males preferentially associate with or avoid certain females'? The MBFB method can be modified to consider such questions.

Total-matrix and within-period permutation methods. In these cases, flips are only made within the 'to' class of individuals; the females in the putative example (see Table 4).

Between-period permutation method. In this case, flips are only carried out within the part of the 1:0 association matrix that represents associations between the two classes; the males–females section in the putative example (see Table 5).

When testing for between-class preferences, using any of the three permutation methods, test statistics are only calculated for the association indexes between individuals of the two classes.

Tests for differences in overall gregariousness

The MBFB method and its modifications test for preferred or avoided associations. However, the question of whether there are differences in overall gregariousness between animals is also interesting: perhaps some individuals are found in consistently larger groups than others. We know of no published test for differences in overall gregariousness among individuals, although variation in gregariousness among classes has been examined by Underwood (1981) and Pepper et al. (1999) using variants of a measure often called the typical group size. The typical group size of an animal is the mean group size that it experiences (Jarman 1974). The MBFB method, and its within-period permutation modification, allow variation in typical groups sizes among individuals to be tested. The test statistic, then, is:

$$\text{SD} \left[\frac{\sum_g \left[\frac{x(g,i) \sum_j x(g,j)}{\sum_g x(g,i)} \right]}{\sum_g x(g,i)} \right]$$

where $x(g,i) = 1$ if animal i was a member of group g , and $x(g,i) = 0$ if animal i was not a member of group g . Significantly large values of the test statistic indicate that some animals are consistently found in particularly large groups and others in particularly small groups.

In the case of the between-period permutation method, there is no simple extension to examine individual differences in gregariousness, as the number of associates of each individual in each sampling period is constrained.

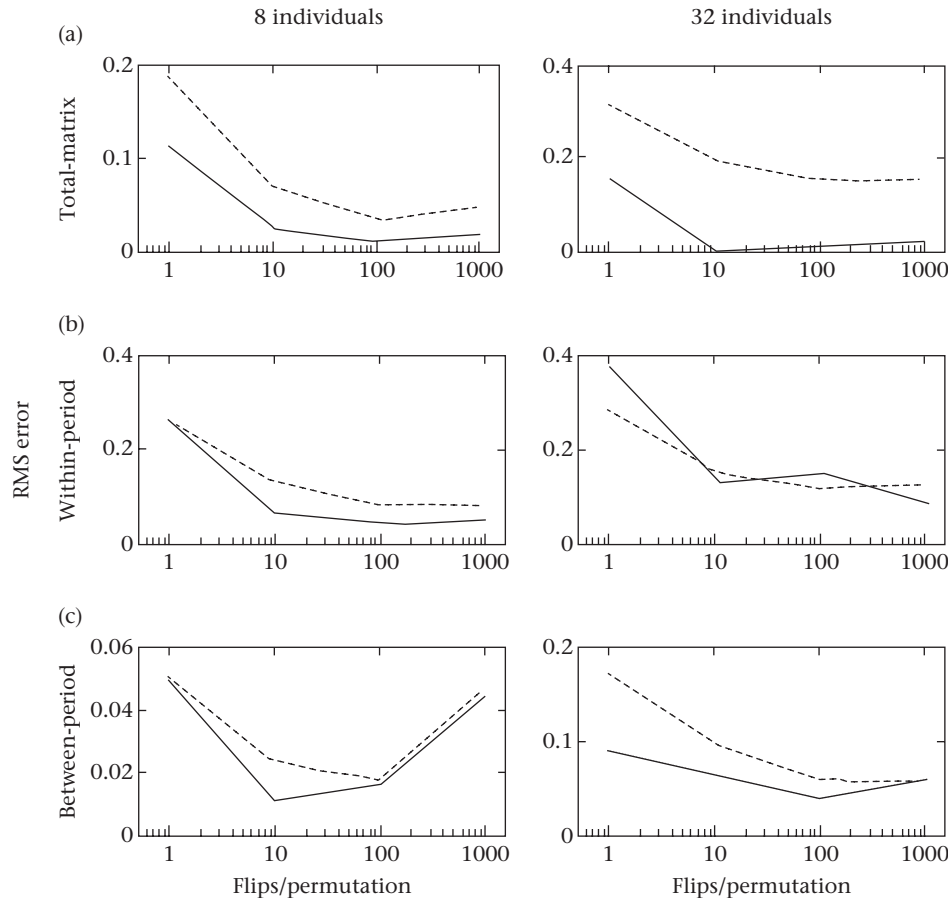


Figure 1. Root mean square (RMS) errors in overall (straight line; standard deviation of association indexes) and dyadic (dashed line) P values for six simulated data sets with the number of flips per permutation (n), for populations of eight individuals (left) or 32 individuals (right), and using either the total-matrix permutation method (a), within-sample permutation method (b), or between-sample permutation method (c). In all cases there were 200 sampling periods, mean group size was 3 (i.e. probability of an individual being in a particular group was 3/population size), and the simple-ratio association index (Cairns & Schwager 1987) was used. With populations of eight individuals, each run took $t = 100$ s, there were 10 runs for each combination of the number of flips and permutation method, and there were two groups per sampling period (except for the within-sample permutation method when there were 10 groups per sampling period). With populations of 32 individuals, each run took $t = 500$ s, there were five runs for each combination of the number of flips and permutation method, and there were 16 groups per sampling period. RMS errors were calculated by comparing the P value(s) for any run with 'baseline' P values from running the routine with $n = 100$ flips per permutation over 10 000 s (with eight individuals) or 50 000 s (with 32 individuals).

However, random distributions of numbers of associates could be generated by permuting identities of the individuals in each sampling period (i.e. randomly permuting the name labels in Table 3 as in the Mantel test; Schnell et al. 1985), and examining the distribution of a test statistic such as the standard deviation of the mean number of associates of each individual over sampling periods during which it was identified. We have not tried this possibility.

These techniques of examining differences in overall gregariousness have some of the benefits of the MBFB methods. P values can be calculated for each individual (the proportion of permutations with typical group size for that individual greater than the real typical group size for that individual; cf. dyadic P values), to identify individuals with significantly large or small typical group sizes. Another potential extension is to examine variation in gregariousness between classes of individuals, so that the tested hypothesis is something like: 'do males differ in the number of females that they associate with?'

Conclusion

Both our own use of these permutation tests to examine preferred/avoided associations, and our reading of their usage by others, have confirmed that they are important tools in the analysis of social structure and that they are flexible, but they have to be used carefully. Users should carefully consider the most appropriate permutation method, the test statistic(s) to be used, and should ensure that enough permutations have been carried out to stabilize P values. Two particularly important issues that have come to light are the misuse of dyadic P values as measures of the strength of association; and the frequent need for considerably more permutations to stabilize dyadic P values than to stabilize overall P values.

In carrying out these tests, several more general issues also need careful consideration. These include the choice of association index and whether or not to correct P values for multiple comparisons if dyadic associations are being

Table 4. Testing for preferred or avoided associations between classes 'M' and 'F' using the total-matrix or within-period method: flips are made only within the F class of group-individual matrix

Group	Individual/Class						
	A	B	C	D	E	F	G
	M	M	M	F	F	F	F
a	1	1	0	0	0	0	0
b	1	1	1	1	0	0	0
c	1	1	0	0	0	0	0
d	1	1	1	1	0	0	0
e	0	0	1	0	1	1	1
f	0	0	1	0	0	1	1
g	0	0	0	0	1	1	1
h	0	0	0	0	0	1	1
i	0	1	0	0	0	0	1
j	0	1	0	0	0	1	1
k	0	1	0	0	0	1	0
l	1	0	1	0	0	0	0
m	0	0	1	0	1	0	0
n	1	0	0	0	1	0	0
o	1	0	0	0	1	0	0
p	0	0	1	1	0	0	0
q	1	0	0	1	0	0	0
r	1	0	1	1	0	0	0
s	1	0	1	1	1	0	0
t	0	0	1	0	1	0	0

tested. Cairns & Schwager (1987), Ginsberg & Young (1992) and others consider the choice of association index; we have also analysed data sets sequentially using two or more association indexes to investigate the robustness of the results with respect to choice of association index. Even when there are moderate numbers of individuals in the study population, many dyads may be being tested, and so correcting for multiple comparisons will have a major impact on the power of the tests. For instance, with 20 individuals, there are 190 possible dyads, and the critical P value using the Bonferroni procedure for an α of 0.05 becomes 0.0003.

Table 5. Testing for preferred or avoided associations between classes 'M' and 'F' using the between-period permutation method: flips are made, and test statistics calculated, only within the bold areas of the association matrix

	Individual/Class								
	A	B	C	D	E	F	G	H	
	M	M	M	F	F	F	F	F	
A	M	—	1	0	0	0	1	0	1
B	M	1	—	1	1	0	0	1	1
C	M	0	1	—	1	1	1	1	0
D	F	0	1	1	—	0	1	0	0
E	F	0	0	1	0	—	0	1	1
F	F	1	0	1	1	0	—	0	0
G	F	0	1	1	0	1	0	—	1
H	F	1	1	0	0	1	0	1	—

In this paper we have described two potentially useful modifications of the original methods, but more are possible (see Whitehead 1999 for additional ideas).

This research was funded by the Natural Sciences and Engineering Research Council of Canada. Thanks to Bruce Smith for statistical advice and to John Pepper for his most constructive review of the manuscript.

References

- Bejder, L., Fletcher, D. & Bräger, S. 1998. A method for testing association patterns of social animals. *Animal Behaviour*, **56**, 719–725.
- Cairns, S. J. & Schwager, S. J. 1987. A comparison of association indices. *Animal Behaviour*, **35**, 1454–1469.
- Christal, J. & Whitehead, H. 2001. Social affiliations within sperm whale (*Physeter macrocephalus*) groups. *Ethology*, **107**, 323–340.
- Ginsberg, J. R. & Young, T. P. 1992. Measuring association between individuals or groups in behavioural studies. *Animal Behaviour*, **44**, 377–379.
- Hemelrijk, C. K. 1990. Models of, and tests for, reciprocity, unidirectionality and other social interaction patterns at a group level. *Animal Behaviour*, **39**, 1013–1029.
- Jarman, P. J. 1974. The social organization of antelope in relation to their ecology. *Behaviour*, **48**, 215–267.
- Johnson, D. H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management*, **63**, 763–772.
- Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E. & Dawson, S. M. 2003. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this trait? *Behavioral Ecology and Sociobiology*, **54**, 396–405.
- Manly, B. F. J. 1995. A note on the analysis of species co-occurrences. *Ecology*, **76**, 1109–1115.
- Manly, B. F. J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd edn. London: Chapman & Hall.
- Morgan, B. J. T., Simpson, M. J. A., Hanby, J. P. & Hall-Craggs, J. 1976. Visualizing interaction and sequential data in animal behaviour: theory and application of cluster-analysis methods. *Behaviour*, **56**, 1–43.
- Pepper, J. W., Mitani, J. C. & Watts, D. P. 1999. General gregariousness and specific social preferences among wild chimpanzees. *International Journal of Primatology*, **20**, 613–632.
- Schnell, G. D., Watt, D. J. & Douglas, M. E. 1985. Statistical comparison of proximity matrices: applications in animal behaviour. *Animal Behaviour*, **33**, 239–253.
- Slooten, E., Dawson, S. M. & Whitehead, H. 1993. Associations among photographically identified Hector's dolphins. *Canadian Journal of Zoology*, **71**, 2311–2318.
- Smolker, R. A., Richards, A. F., Connor, R. C. & Pepper, J. W. 1992. Sex differences in patterns of association among Indian Ocean bottlenose dolphins. *Behaviour*, **123**, 38–69.
- Underwood, R. 1981. Companion preference in an eland herd. *African Journal of Ecology*, **19**, 341–354.
- Whitehead, H. 1999. Testing association patterns of social animals. *Animal Behaviour*, **57**, F26–29.
- Whitehead, H. & Dufault, S. 1999. Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Advances in the Study of Behavior*, **28**, 33–74.
- Whitehead, H., Silver, R. & Harcourt, P. 1982. The migration of humpback whales along the northeast coast of Newfoundland. *Canadian Journal of Zoology*, **60**, 2173–2179.