SOME ASPECTS OF RESEARCH DESIGN AND THEIR IMPLICATIONS IN THE OBSERVATIONAL STUDY OF BEHAVIOUR

by

R. I. M. DUNBAR[1])
(Department of Psychology, University of Bristol, England)

(Acc. 5-VIII-1975)

INTRODUCTION

The success of any research project depends to a large extent on the amount of forethought devoted to decision-making before the observer even begins to look at his animals. In any given case, there are five basic decisions which the investigator has to make, whether he does so consciously or not (J. ALTMANN, 1974). These may be summarised as follows:

1. The problem to be investigated has to be defined;
2. The kinds of behavioural parameters most appropriate to answering that particular question must be identified;
3. A sampling strategy which will provide an unbiassed estimate of these parameters must be chosen;
4. A method which is both suitable for recording the data and practicable in the field must be selected;
5. The most appropriate statistical test(s) for analysing the data in the form obtained must be chosen.

Each of these decisions poses its own problems and difficulties. Furthermore, the decision reached at any one of these points places constraints on the options which are open to the investigator at other points. It is well known, for instance, that an inappropriate choice of statistical test can invalidate the conclusions drawn from a set of data. It goes without saying, therefore, that the data must be recorded in a form which is suitable for

1) I am grateful to the many people with whom I have discussed the problems which have been the subject of this paper. I am especially indebted to Mrs Jeanne ALTMANN, Dr J. H. CROOK, Professor R. A. HINDE, Jenny INGRAM and Dr E. M. RUSSELL for their comments on an earlier draft of the manuscript; their criticisms greatly helped to clarify many of the issues. The field work was made possible by grants from the Science Research Council, London, and the Wenner-Gren Foundation for Anthropological Research while the author was in receipt of an S. R. C. Overseas Scholarship.

analysis by an available test. It is perhaps, less often appreciated, however, that inappropriate solutions at any of the other stages may have equally unfortunate results (see, for instance, the discussion of sampling methods by J. ALTMANN, 1974).

The increasing emphasis placed on the quantification of data in recent years has forced the observer to consider these five points of research design more seriously than has formerly been the case. This is reflected in the number of papers recently published which have been concerned purely with research methodology (*e.g.* HUTT & HUTT, 1970; HINDE, 1973; J. ALTMANN, 1974). Unlike these latter works, the present paper is not intended to be an observer's guide: rather, I aim simply to draw attention to some important aspects of research design which have not been discussed in any detail elsewhere. Whereas most methodological papers have been concerned with points (3) and (4) above, I shall be concerned primarily with point (2), that is, the choice of behavioural parameters which are relevant to the research problem. This should not, however, be taken to imply that this aspect of research design is more important than the others.

The particular problem that prompted this paper was how best to quantify social relationships among groups of monkeys (in this case gelada baboons, *Theropithecus gelada*). However, although I shall be concerned mainly with the effects of using different parameters to quantify social relationships, many of the conclusions drawn apply to other areas of behavioural research.

1. THE RESEARCH PROBLEM: RATIONALE AND AIMS

Many field workers find it useful to describe the overall pattern of social relationships in the group of animals under study. By doing so, the observer can identify social subgroups, the composition and organisation of which can be related to the structure and dynamics of the group as a whole. Thus, rather than treating dominance hierarchies *in vacuo*, we may be able to determine how dominance relationships relate to preferences in social partners. Do, for instance, animals that frequently engage in "friendly" social interaction support each other when attacked by other individuals?

In studying social relationships as such, we commonly suppose that we are investigating something which is relevant to the animals themselves. Since, on the whole, animals tend not to interact at random, we assume that when they do interact with a particular individual, in some real sense they choose to do so. In other words, we assume that the interaction concerned is not some coincidental result of a random series of events. Some authors have termed these relationships "social affinities" or "social bonds" (see, for instance, KUMMER, 1968; ALDRICH-BLAKE *et al.*, 1971; SADE, 1973; DUNBAR & DUNBAR, 1975). In general, these terms are used in an *operational* sense: that is to say, a social bond is defined in terms of a particularly high rate of interaction between two individuals. Naturally, we can say little about what the animals themselves (to put it anthropomorphically) "think" since we are restricted to considering only the observable aspects of their behaviour. Nonetheless, many authors have found a description of the "communication network" within the group (*i.e.* the channels or routes of most frequent communication) to be heuristically useful insofar as it provides a quantitative, first level approach to identifying the underlying structure of the system (*e.g.* S. ALTMANN, 1968; KUMMER, 1968; SADE, 1973).

However, different field workers have used different quantitative methods for determining such relationships. Consider the following few examples: KUMMER (1968) in his study of *Papio hamadryas* used the number of 1 min intervals in which individuals interacted; SADE (1965, 1973) used the frequencies with which individuals groomed each other in his study of *Macaca mulatta*, while S. ALTMANN (1968), in studying the relationships between age-sex classes in this species, used the frequencies with which social signals (or acts) were exchanged; LESKES & ACHESON (1971) used the frequencies with which individual *Colobus guereza* exchanged social acts; and in our own study of *Theropithecus gelada* we used the number of 2 min instantaneous scan-samples in which individuals were interacting as well as the frequencies with which age-sex classes interacted (DUNBAR & DUNBAR, 1975).

On the other hand, some authors have used spatial criteria, an approach originally suggested by CARPENTER (1964) who argued that the metric distance between any two individuals could be used as an index of the strength of the "bond" between them. I list a few examples of the bewildering array of spatial measures which have been used. KUMMER (1968) used the metric distance between two individuals as well as relative spatial relationships (his "nearest-neighbour" method); DEAG (1974) also used a "nearest-neighbour" method in his study of *Macaca sylvanus*; CHALMERS (1968) recorded the frequencies with which *Cercocebus albigena* age-sex classes were within 3 ft of each other, while BRAMBLETT (1972), in a captivity study of *T. gelada*, recorded the number of time intervals in which individuals were within 10 ft of each other; STRUHSAKER (1967) recorded the frequencies with which *Cercopithecus aethiops* individuals spent the night in the same sleeping subgroups, and DUNBAR & NATHAN (1972) adopted a comparable strategy in the case of *Papio papio*; ALDRICH-BLAKE *et al.* (1971) analysed the frequencies with which *Papio anubis* individuals followed each other during linear progressions.

All these different procedures are ostensibly measuring the same thing, namely the strengths of the social relationships between different individuals or classes of individuals. All, explicitly or implicitly, assume that individuals who interact (or sit together) often have a stronger social relationship or "bond" than individuals who do so only rarely. While the assumption that frequent interaction reflects a strong "social bond" in the psychological sense remains open to question, we can at least assert that some animals interact or sit together more often than do others.

However, we may justifiably ask whether the different measures which have been used provide data which are comparable in form. Do animals who sit together often necessarily interact with each other or groom each other often? What difference would it have made to the resulting picture of social relationships if the observer in question had used a different measure? There is some indication in the literature that different measures do produce results which are sometimes radically different. KUMMER (1968), for example, found that patterns of interaction did not necessarily correspond to patterns of spatial proximity. Likewise, SIMPSON (1973) found that different measures of the distribution of grooming among adult male chimpanzees did not necessarily correlate. This raises the critical question as to whether we can legitimately compare data for different species if they were obtained in different ways.

Our purpose, then, will be to examine the effect of using different behavioural parameters and sampling methods to estimate social relationships among a set of animals. Naturally, time and space preclude an attempt to investigate all possible methods, and I have chosen those which are either commonly used or representative of a group of methods. My aim is not to show that any one procedure is absolutely better than any other, but simply to illustrate the effects of choosing one particular approach and to try to show why this is so.

2. BEHAVIOURAL PARAMETERS

As J. ALTMANN (1974) has pointed out, the way in which the research question is phrased both reflects the kinds of problems in which the observer is interested and determines to a large extent the kinds of data which will be relevant to that particular question. This in turn may restrict the ways in which the behaviour in question can be sampled. Our present concern requires that we have as many different measures as possible in order to be able to compare them. I have selected three types of relationship to examine.

These are (a) the extent to which animals interact with each other, (b) the extent to which they groom each other and (c) the extent to which they are in a particular spatial relationship. Any given behaviour can be quantified in two quite different ways, namely, either by counting events (frequency or rate measures) or by scoring states as a proportion of the total sample time (durational measures) and these are by no means equivalent (see J. ALTMANN, 1974). Each of the three behavioural parameters listed above can be quantified in terms of both frequency and duration. For convenience, I shall postpone the discussion of the actual measures used to the following section. Before doing so, however, the kinds of behaviour which are to be sampled in each case must be specified.

In the case of (a) above, only non-agonistic "associative" behaviours will be analysed. Interactions which are agonistic in character will be excluded from consideration on the grounds that they reflect the existence of "negative" relationships, although some authors have not made this distinction. In practice, agonistic interactions are rare compared with non-agonistic interactions in most species and the amount of distortion due to their inclusion may be minimal. However, when dealing with individuals rather than age-sex classes, there is an increased risk of finding individuals who interact relatively frequently but always in an agonistic manner. To include agonistic interactions might lead to the conclusion that there was a strong social relationship between the individuals concerned. Clearly this would be erroneous since the relationship between them is qualitatively quite different to that between two individuals who only groom each other, even though they interact less frequently.

I do not distinguish between the different kinds of non-agonistic interaction (*e.g.* grooming, play, sex, *etc.*) on the grounds that, at this stage, there is no a priori reason to suppose that social relationships are exclusively expressed in terms of a particular behaviour. To assume that the relative strengths of social relationships are reflected in the extent to which individuals groom each other would bias against those individuals or classes of individuals, such as juveniles, who normally play rather than groom. This is not to say that grooming relationships themselves might not be of interest, but the analysis of particular types of interaction should be considered a second order approach once the general outlines have been established. I have, however, included grooming interactions (b above) as a separate measure, partly because they are commonly used as an index of social relationships and partly because grooming is often said to be the most important form of "friendly" behaviour among primates.

With regard to the spatial measure (c above), we shall record the extent to which animals are within arm's reach of each other. The usual rationale for this approach is that, since the animals do not normally mix at random, the fact that two individuals are spatially close reflects the existence of some "attraction" between them. By implication, therefore, it is necessary to ensure that data of this kind are not collected at times when there is reason to believe that this assumption does not hold. Such circumstances are likely to include occasions when the animals are moving rapidly (especially if fleeing from potential danger) and when the group is highly disturbed by agonistic encounters of a particularly severe kind. Under such conditions, the animals may be assorted more or less at random since their main concern may be to get away from the source of the disturbance rather than whom they are next to. Both KUMMER (1968) and DEAG (1974), for instance, recorded data of this kind only from resting troops, and in the present case a similar condition was imposed.

3. SAMPLING STRATEGIES

Sampling methods have been discussed in detail by J. ALTMANN (1974) and I shall here simply describe the various strategies used.

In all, seven different measures or estimators of social relationships have been taken. These result from different combinations of behavioural parameters and sampling methods and are as follows:

(1) The number of social contacts between individuals (*i.e.* the number of times they interacted, irrespective of the number of social acts exchanged in each interaction);

(2) The number of social acts exchanged between individuals;

(3) The number of 30 sec time intervals in which individuals interacted (One-Zero sampling);

(4) The number of 60 sec instantaneous or point samples in which individuals were interacting;

(5) The number of grooming bouts exchanged between individuals;

(6) The number of 60 sec instantaneous or point samples in which individuals were grooming;

(7) The number of 30 sec time intervals in which individuals were within arm's reach of each other (One-Zero sampling).

The sampling strategies and some of the terms used require explanation and definition. A "social contact" is defined as the exchange of an uninterrupted series of social acts between two individuals, being defined as terminating when the individuals concerned cease to exchange such acts and do something else (*e.g.* feed, move apart, interact with another animal, *etc.*). In strategy (1), we simply count the number of such contacts that occur and ignore the number of social acts exchanged in any given case. By definition, only interactions involving the exchange of non-agonistic signals will be considered.

A "social act" refers to any element of the species' signal repertoire when used in a "face-to-face" interaction between a pair of animals. In strategy (2), we count the number of such acts exchanged between interacting dyads. Individual social acts include grooming, touching, presenting, mounting, the various play behaviours, *etc.* Agonistic acts (threats, aggressive contact behaviours, submissive behaviours) and signals such as alarm and contact calls which are not generally directed at a specific individual are discounted in the present context.

A "grooming bout" is defined as a continuous series of grooming actions by one individual. As in the case of social contacts, a grooming bout is defined as terminating when the groomer stops performing grooming actions and does something else (*e.g.* sits looking around, presents to be groomed, moves away or is groomed by his partner). I include among the actions of grooming not only searching through the fur and removing particles, but also eating things picked from the fur or skin and lipsmacking (see SPARKS, 1967). A more detailed discussion of this is given elsewhere (DUNBAR & DUNBAR, 1975). In strategy (5), we count the number of individual bouts of grooming exchanged between two individuals (irrespective of which animal is the groomer).

An example may help to clarify the differences between these measures.

Suppose the following sequence was recorded from a single subject denoted by animal A: A approaches B and (1) presents his rear to him. B (2) touches A's rear, (3) peers closely at it and then (4) briefly mounts him with (5) vigorous lipsmacking. When B dismounts, A turns and (6) begins to groom B. After grooming for some time, A stops and (7) presents his side to B; then B (8) grooms A. Later, B stops grooming and sits looking around. A moves away a few metres and begins to feed. After feeding for a while, A walks back over to B and (9) begins to groom him. When B stops grooming, A (10) presents his back to B and B (11) grooms A again. A few minutes later, B walks away and A begins to feed.

In this example, two separate social contacts occurred which were separated by a clear break in the exchange of social signals between the animals, namely when A moved away to feed and B sat looking around between (8) and (9). During these two interactions, four separate bouts of grooming were recorded (at points 6, 8, 9 and 11) and a total of 11 social acts were exchanged.

These three strategies (1, 2 and 5) are event frequency (or rate) measures and the data are properly expressed, even if unstated, as the number of events per unit time. The remaining four strategies, on the other hand, are durational measures, the data being strictly expressed in terms of the proportion of time spent in a given state.

In strategies 3 and 7, the sampling procedure is based on One-Zero or interval sampling. The sample period (in this case 30 min) is divided up into a series of equal time intervals (here 30 sec). So long as the relevant relationship occurs within any given 30 sec interval, a "plus" is scored; if the relationship does not occur, a "minus" is scored. Only a single plus is scored for a particular pair of interactees in any given interval, irrespective of how many times they actually interacted (or sat together) in that interval. This sampling strategy has been referred to by a number of terms, but I shall follow J. ALTMANN (1974) in terming it "One-Zero" sampling. (For present purposes I have considered One-Zero sampling to be an estimate of the proportion of time. Some authors have used a One-Zero sampling strategy to estimate event frequencies, but, as ALTMANN & WAGNER (1970) and FIENBERG (1972) have pointed out, the strategy does not give a true frequency estimate and I have refrained from sampling events in this way. For a fuller discussion of this, see, HINDE (1973) and J. ALTMANN (1974).)

In strategies 3 and 6, the data are also recorded in the form of +'s and -'s, but in this case samples take the form of instantaneous samples carried out at predetermined intervals of time (here 60 sec). Every 60 sec we determine whether or not two individuals are interacting in the way specified. This strategy has likewise been referred to by a number of terms, but for simplicity I shall use the term "point" sampling.

METHODS

1. Recording methods

The data were obtained from a set of eleven 30 min samples of individual gelada baboons (*Theropithecus gelada*) recorded for this purpose during our 1971-2 study of this species in the Simien Mountains National Park, Ethiopia. The recording method was so designed that data were simultaneously obtained on each of the seven different measures. In each case, a single subject animal was selected at random from the herd, and his behaviour was recorded on a checksheet graduated into sixty 30 sec intervals. Everything the subject did (or had done to him) and the age-sex class of his interactee was recorded in the interval in which it occurred. Persistent behaviours such as grooming were recorded in each interval in which they occurred, the end of a bout being marked by a vertical line behind the last entered symbol. At 60 sec intervals, an instantaneous sample of what the subject was doing (and, where appropriate, with whom) was made. In addition, the exact times of onset and termination of all grooming bouts were noted to the nearest sec (with a possible error of 2-3 secs) and the age-sex classes of all animals within arm's reach of the subject during any part of each 30 sec inerval were recorded.

Observation distances were of the order of 3-10 m, visibility conditions in all cases being excellent. Inter-observer reliability when recording samples of this kind was exceptionally high (95%: see DUNBAR & DUNBAR, 1975). Eleven age-sex classes were distinguished in this analysis: these are described in full in DUNBAR & DUNBAR (1975).

No attempt was made to sample either from a single known individual or from individuals of the same age-sex class. Although either of these might have been preferable in terms of realism, we aimed to sample a broad spectrum of all age-sex classes except infants. This has the advantage of including in the sample a variety of individuals whose characteristic interactions are likely to differ, hence emphasising any differences which might exist between the measures.

2. Analysis of data.

The major analyses of the data will be in terms of the proportional distribution of social interactions (however measured) among the various age-sex classes of interactee. For these purposes, the eleven samples have been pooled and treated as though they were obtained from the same individual or class of individuals. *It is important to note, therefore, that no conclusions concerning the behaviour of gelada baboons should be drawn from the data presented below.* In analysing the distribution of social interactions, all triadic or multipartite interactions have been treated as a set of dyads.

Spearman rank correlation tests have been used to determine whether the various sampling strategies give the same ordinal ranking to the data. These analyses represent the kind of qualitative and semi-quantitative comparisons that are often made, namely those of the "A-interacts-more-with-B-than-with-C" kind. However, since a large number of comparisons are made, significance levels cannot be attached directly to the resulting r_s values (SIEGEL, 1956). The results of the tests are used simply to show more clearly the extent to which the individual measures agree in their rankings of the data. The overall degree of correlation between the measures is tested for significance using Kendall's coefficient of concordance.

Some direct tests of absolute numerical comparability have been carried out using the Wilcoxon matched-pairs signed-ranks test. In general, however, most numerical comparisons will be made without the use of statistical tests since the differences are generally sufficiently extreme not to require statistical verification. (There is, however, the additional problem that only goodness-of-fit tests are appropriate in some of these cases, but, since independence of the data cannot be assumed in six of the seven measures, such tests cannot be used. Only in the case of the frequency of social contacts measure can independenct be demonstrated: for a discussion of this, see Appendix C of DUNBAR & DUNBAR, 1975).

RESULTS AND DISCUSSION

1. ESTIMATING SOCIAL RELATIONSHIPS

The frequencies with which the subjects (all samples pooled) interacted (or sat with) each of the various age-sex classes as determined by the seven sampling strategies are given in Table 1. In each case, the raw number of units of interaction or association with each class is given. Strictly speaking, estimators 3, 4, 6 and 7 should be interpreted in terms of the proportion of the total number of samples taken (330 in the case of strategies 4 and 6, 660 in the case of strategies 3 and 7). However, for simplicity the raw scores are entered into the appropriate columns rather than the proportion of samples.

As a first step, each pair of measures may be compared to determine whether or not they rank the interactees in the same order. The results of the Spearman rank correlation tests are given in Table 2. It can be seen that,

TABLE 1

Frequencies with which subjects "interacted" with the various age-sex classes as determined in different ways (all subjects combined)

	Age-sex class of interactee											
Estimator	Ad ;	6yr ;	5yr ;	4yr ;	Ad	3yr ;	3yr	2yr ;	2yr	Yearl	Inf	Total
1. Number of contacts	3	3	1	2	8	11	3	3	2	10	1	47
2. Number of acts	19	10	3	7	47	34	8	15	8	17	1	169
3. One-Zero interacting	34	33	3	17	99	49	11	39	9	10	1	305
4. Point interacting	15	15	1	7	44	19	3	16	5	1	–	126
5. Grooming bouts	16	9	1	5	38	18	7	15	2	3	–	114
6. Point grooming	15	15	1	7	44	17	3	16	4	–	–	122
7. Spatial measure	45	37	4	19	111	51	17	40	21	67	3	415

TABLE 2

Comparison of the results given by the various estimators in Table 1 by Spearman rank correlation test, the values of r_s being given in each case

	Contacts	Acts	One-Zero interacting	Point interacting	Grooming bouts	Point grooming
Acts	0.900					
One-Zero interacting	0.774	0.816				
Point interacting	0.862	0.778	0.986			
Grooming bouts	0.753	0.880	0.973	0.932		
Point grooming	0.519	0.710	0.941	0.993	0.909	
Spatial measure	0.888	0.961	0.673	0.685	0.818	0.607

N = 11 in each case.

on the whole, the correlations between the measures are quite high. Of the twenty-one pairings, thirteen correlate at better than $r_s = 0.800$, while in only one case is the value of r_s below 0.600.

Such a generally high level of agreement between the seven measures might be taken to imply that, at least in the case of the species under study, which strategy is used to estimate overall social relationships is relatively immaterial so long as the data are ordinally scaled. Indeed, calculation of Kendall's coefficient of concordance between the seven measures gives W = 0.840, which indicates a significant degree of agreement between the measures ($\chi^2 = 58.8$, $p<0.001$, df = 10).

However, the fact that in many cases the correlations are far from perfect raises the issue of the numerical comparability of the various measures. Consideration of the data in Table 1 shows that although each estimator places adult females, 3 year males, 2 year males, adult males and 6 year males in the top six ranks, they do not do so in exactly the same order. Furthermore, the class of yearlings is ranked markedly differently by the various measures, and it can be seen that this class accounts for much of the discrepancy in many cases.

These discrepancies imply that, although the measures co-vary, they do not do so monotonically. This can readily be demonstrated by converting the raw scores to percentages of the total number of samples in each case. These are given in Table 3 in the form of probabilities, each row summing to unity.

TABLE 3

Relative strength of relationship with each age-sex class as determined by the different estimators given as a probability (based on data in Table 1)

	Age-sex class of interactee										
Estimator	Ad ♂	6yr ♂	5yr ♂	4yr ♂	Ad ♀	3yr ♂	3yr ♀	2yr ♂	2yr ♀	Yearl	In…
1. Number of contacts	.064	.064	.021	.043	.170	.234	.064	.064	.043	.213	.02
2. Number of acts	.112	.059	.018	.041	.278	.201	.047	.089	.047	.101	.00
3. One-Zero interacting	.111	.108	.010	.056	.325	.161	.036	.128	.030	.033	.00
4. Point interacting	.119	.110	.008	.056	.349	.151	.024	.127	.040	.008	.00
5. Grooming bouts	.140	.079	.009	.044	.333	.158	.061	.132	.018	.026	.000
6. Point grooming	.123	.123	.008	.057	.361	.139	.025	.131	.033	.000	.000
7. Spatial measure	.108	.080	.010	.016	.267	.123	.011	.006	.051	.161	.007

(In the case of the four durational measures, there is a major difficulty in this respect. In order to transform the raw data into a form comparable to that for the three frequency measures, the total number of samples taken cannot be used as the denominator since the rows would not then sum to unity. On the other hand, since a subject could be interacting or sitting with more than one animal during any given sample, the number of intervals or point samples in which he was interacting cannot be used either, since the row totals would then be greater than unity. It has therefore been necessary to use the sum of the appropriate row in Table 1 as the denominator in each case. This raises some difficulty over the precise interpretation of the probabilities, since they no longer represent either the proportion of time or the proportion of interaction time that the subject spent interacting. Perhaps the best interpretation to place on these probabilities is one in terms of the proportional distribution of the subject's social contacts or spatial associations.)

Consideration of these data clearly illustrates the different pictures of gelada social relationships which would result from use of the different measures. Frequencies of social contact, for instance, would suggest that 6.4% of the subject's social activity was with adult males and 21.3% with yearlings. On the other hand, the number of acts exchanged would give values of 11.2% and 10.1% respectively, while the number of grooming bouts exchanged would give values of 14.0% and 2.6%. Even more extreme would be the number of point samples spent interacting: 11.9% as against 0.8%.

Hence, although we might be able to compare different animal's social relationships using ordinally scaled data with a reasonably high expectation of concordance between different measures, we cannot necessarily make a more direct comparison between the relative probabilities of interaction.

2. DIFFERENCES BETWEEN MEASURES

These differences draw attention to one important aspect of social relationships. In a complex social system, "social bonds" (in the psychological sense) may be expressed in a variety of different ways. Hence, whichever index we select to estimate the strengths of these relationships will provide data which reflect only certain facets of this system.

The mean number of acts exchanged per social contact, for instance, is not the same for all classes of interactee. Interactions with some classes average five or six acts per contact, while those with others average only one or two (see Table 1). Hence, by taking frequencies of social contact as an index of the strength of the relationships between individuals, the data are biassed in favour of those classes who exchange fewest acts per social contact, while the reverse is the case if the number of acts exchanged is used as an index. Likewise, consideration only of grooming bouts will bias the results against those individuals or classes among whom grooming is a rare form of social expression. Conversely, it would tend to overemphasise relationships between classes who do little else but groom when they do interact, even though they may not be especially sociable.

Again, if we record the proportion of time spent interacting, we may find that similar biasses emerge due to the fact that interactions with some classes tend to be of shorter duration than those with others. Not only may these measures differ from the distribution of social contacts, but they may also differ from the distribution of acts if animals tend to exchange different kinds of signals in their interactions with different individuals.

The strongest correlations obtained were between the three estimators which measured the proportion of time individuals spent interacting and grooming (strategies 3, 4 and 6). This can be attributed to the fact that, in the present case, a large proportion of the social acts exchanged were grooming (see Table 1). Since grooming tends to account for a great deal more time than other behaviours, it is not too surprising to find that a measure of the proportion of time spent interacting correlates well with the proportion of time spent grooming. This may not, however, necessarily be so in all cases, since the interactions of some classes, such as juveniles, tend to involve a great deal more play than grooming, particularly when they are interacting with peers. In this case, the assumption that grooming provides an accurate estimate of the strengths of social relationships may be seriously misleading.

Similarly, a measure of spatial association may lead to a rather different picture of the relationships among a group of animals than an interactional measure would give. The reason for this is fairly obvious: an animal may spend a great deal of time close to certain individuals, but interact with them only rarely. This could be due to one of several reasons. Infants, for instance, might often be found near adult males because their mothers commonly associate with this age-sex class: nonetheless, infants might interact with adult males only rarely. The assumption that there is any kind of "social bond" between adult males and infants purely on the basis of frequent spatial proximity would be erroneous, since the apparent relationship is in reality a consequence of the strong relationship between males and females. On the other hand, some animals may actively seek out the proximity of certain individuals, but be unable to interact with them as often as they might wish. KUMMER (1968) found such a relationship between juvenile females and adult males in *Papio hamadryas*, and similar results were obtained in the case of gelada baboons (DUNBAR & DUNBAR, 1975). Thus, in this case, spatial proximity might reflect the juvenile's choice of preferred social partner rather more closely than frequencies of interaction. A third possible interpretation might be that spatial proximity as a whole is irrelevant. Animals might mix more or less at random when not interacting, moving through the group to seek out their preferred social partners as and when they wished to interact. There was evidence to this effect among certain age-sex classes of gelada (DUNBAR & DUNBAR, 1975). In this situation, spatial proximity would be a useful index of social preferences only when the animals were actually interacting. It remains an untested possibility that in many species neighbourhood relationships among a group of animals change according to the nature of the predominant activity. It is not unreasonable to suppose, for instance, that an animal's neighbours might approach a random distribution more often when the group is feeding than when it is socialising.

5. DIFFERENCES BETWEEN FREQUENCY AND DURATION

The differences between frequency and duration can be illustrated more clearly by examining the relationship between the frequency of grooming bouts and the amount of time spent grooming. These data are given for each subject in Table 4, together with the mean duration of grooming bouts in each case.

The number of grooming bouts correlates significantly with the total amount of time spent grooming ($r_s = 0.872$, $p < 0.01$, $N = 11$) even though there is a considerable range in the mean duration of the various subjects' grooming bouts (14.0-124.6 secs). In part, this correlation can be attributed to the fact that, although grooming bouts may vary from a few seconds to as much as 6 min in length, they tend to have a modal distribution of around 70-80 secs (unpublished data). The relationship between the frequency and duration of grooming is, however, by no means linear. Some individuals may

TABLE 4

Comparison of the number of seconds each subject spent grooming with the number of grooming bouts he performed

	Subject										
	1	2	3	4	5	6	7	8	9	10	11
Number of seconds	1050	717	703	674	623	469	299	241	110	20	14
Number of bouts	11	7	9	7	5	9	6	5	3	1	1
Mean bout length (in secs)	95.5	102.4	78.1	96.3	124.6	52.1	49.8	48.2	36.7	20.0	14.0

groom frequently but not for any length of time, whereas others groom infrequently but for a considerable period of time on each occasion (compare subjects 5 and 6 in Table 4). Indeed, there is no consistent relationship between the frequency with which subjects groomed and the mean duration of their grooming bouts ($r_s = 0.596$, $p > 0.10$, two-tailed, $N = 11$), indicating that the most frequent groomers are not necessarily the most persistent. Not only this, but animals may also differ in the length of time for which they are prepared to groom different individuals. SIMPSON (1973), for instance, found this to be the case among adult male chimpanzees.

Consequently, when choosing an appropriate measure for estimating the extent to which individuals interact, we should bear in mind the differences between frequency and duration. In an absolute sense, neither can be considered better than the other, since each examines a different aspect of the situation, although in any given set of circumstances one may be more appropriate in terms of the investigator's interests and the requirements of the research question. It may often be valuable, however, to record both kinds of data.

4. DIFFERENCES BETWEEN ONE-ZERO AND POINT SAMPLING

Although I do not intend to discuss the merits and demerits of the various sampling methods in any detail, it does seem worthwhile examining some of the disadvantages of One-Zero sampling insofar as we have data available to do this.

Throughout the preceeding analyses, I have used One-Zero sampling as a durational measure. Strictly speaking, One-Zero sampling is so designed that it incorporates features of both frequency and duration. Ostensibly, the strategy would seem to have much to recommend it since it may permit the observer to record event frequencies while at the same time obtaining data on duration. The number of studies which have used this particular method attests to this. Regrettably, however, the method leaves much to be desired since considerable error may be incorporated into one or both kinds of data. A more detailed discussion of One-Zero sampling is given by J. ALTMANN (1974). In the present context, I shall consider the differences between One-Zero and point sampling in the extent to which they provide an accurate estimate of the proportion of time spent doing something. I shall again use the proportion of time the subjects spent grooming. That I consider only this aspect of One-Zero sampling should not be taken to imply that I consider it a suitable strategy for recording event frequencies. I have already remarked that this strategy does not give a true estimate of the frequency of event occurrence (see also ALTMANN & WAGNER, 1970; HINDE, 1973; J. ALTMANN, 1974).

Table 5 gives the proportion of time each subject spent grooming based on the actual number of seconds (from Table 4), together with estimates of this obtained by One-Zero and point sampling with different time bases. In the final column is given the overall proportion of time spent grooming by all subjects combined.

Consideration of the data shows that point sampling provides a consistently more accurate estimate of the true function than does One-Zero sampling, irrespective of the length of the sample interval. Comparison of the true function and the estimated function in each case using the Wilcoxon matched-pairs signed-ranks test shows that none of the point sample estimates differ significantly from the true function (T not below 21.0, $p>0.05$ in each case), whereas all the One-Zero samples are significantly different ($T = 0$, $p<0.01$ in each case). Although point samples might differ somewhat from the true value in individual cases, when the sample size is reasonably large the errors tend to cancel each other out and a reasonably close approximation to the true function is obtained. In the case of One-Zero sampling, however, the true function is always over-estimated,

TABLE 5

Estimates of the proportion of time individual subjects spent grooming (based on One-Zero and point sampling with different time bases) compared with the actual proportion of time spent grooming

Estimate	Subject 1	2	3	4	5	6	7	8	9	10	11	Overall %
Number of seconds	58.3	39.8	39.1	37.4	34.6	26.1	16.6	13.4	6.1	1.1	0.8	24.8
One-Zero sampling based on intervals of:												
5 sec	63.1	41.4	40.8	38.9	35.6	28.1	18.1	14.4	6.9	1.3	1.1	26.3
15 sec	69.2	45.0	43.3	42.5	37.5	33.3	21.7	16.7	8.3	1.7	0.8	29.1
30 sec	75.0	51.7	46.6	45.0	40.0	38.3	26.7	20.0	10.0	3.3	1.7	32.6
60 sec	80.0	60.0	56.7	50.0	50.0	43.3	30.0	23.3	16.7	3.3	3.3	37.9
120 sec	86.7	66.7	73.3	53.3	53.3	66.7	46.7	26.7	20.0	6.7	6.7	46.1
Point sampling based on intervals of:												
15 sec	67.5	39.2	38.3	37.5	34.2	25.8	17.5	13.3	5.8	0.8	0.8	25.5
30 sec	60.0	41.7	35.0	36.7	33.3	25.0	16.7	11.7	5.0	1.7	0.0	24.2
60 sec	53.3	43.3	36.7	36.7	36.7	20.0	10.0	13.3	6.7	0.0	0.0	23.9
120 sec	46.7	46.7	46.7	33.3	40.0	26.7	13.3	13.3	6.7	0.0	0.0	24.8

the extent of the over-estimation becoming greater as the length of the sample interval increases. One-Zero sampling yields a reasonably accurate estimate only when the sample interval is very small relative to the typical duration of the behaviour in question.

The cause of this is not difficult to see. In One-Zero sampling, we record simply whether or not the behaviour in question occurs, irrespective of how much of the interval it actually occupies. Since the moments of onset and termination of any given bout are most likely to occur in the middle of an interval, each bout will be credited with lasting up to two full time units longer than it did. Clearly, the longer the length of the sample interval, the greater will be the over-estimation. The extreme case will occur when the sample interval is the same as the sample period (in the present case 30 min). An animal who spent only 10 sec grooming during this period would in this case be credited with spending 100% of his time grooming, whereas in fact he spent less than 0.6%. It can readily be seen that an accurate estimate of the proportion of time spent grooming (or indeed anything else) can only be obtained when the sample interval approximates the actual unit of measurement (*i.e.* one second). In this event, it would obviously be a great deal easier to record transition times and calculate the durations from these. As J. ALTMANN (1974) points out, One-Zero sampling can only provide an upper limit on the proportion of time accounted for by a given state. As we have seen, this upper limit may often be absurdly high: at best, it will simply be misleading.

This may be particularly critical when comparing species or individuals who differ in the mean duration of their grooming bouts. We cannot necessarily assume that the extent of the over-estimation resulting from the use of One-Zero sampling will be consistent in each case. This can be seen from Table 5. Although, for instance, the absolute difference between the true function and the 30 sec One-Zero estimate correlates significantly with the actual amount of time spent grooming ($r_s = 0.791$, $p<0.05$, two-tailed, $N = 11$), the correlation is by no means perfect. To a large extent, this is due to the fact that the mean duration of individual subject's grooming bouts varies considerably (Table 4). The discrepancy between the true function and the estimated function is least when the subject's mean grooming bout length approximates some multiple of the sample interval (*i.e.* 30, 60, 90 or 120 secs; see subjects 5, 8 and 9 in Table 4) and is greatest when the mean duration differs most from these values (subjects 1, 2, 6 and 7).

This problem does not arise in the case of point sampling since this strategy is based on instantaneous sampling. Hence, the probability that the activity in question will be recorded in a given sample will depend simply on the proportion of time it actually occupies. A behaviour pattern which accounts for 30% of an animal's time will have a probability of being recorded in a sample of precisely 0.3. Hence, point sampling is likely to give a consistently reliable estimate of the true function, and this will be so whatever the length of the inter-trial interval. On balance, therefore, a point sampling strategy will always be superior to One-Zero sampling. It is also considerably easier to use, both in the case of individual subjects and when large numbers of animals have to be sampled simultaneously.

CONCLUSIONS

I observed at the outset of this paper that each of the five decisions which the investigator has to make places constraints on the options that are open to him at each of the remaining decision points. Normally, we make the decisions in the order in which they were listed on p. 78. In other words, we begin, not unnaturally, by posing a particular problem to investigate, then we decide what kinds of behaviours are relevant to it, choose a suitable sampling strategy, decide how to record the data in the field and finally (usually after the data have been collected) we wonder how the results can be analysed. As we progress through these decisions, the number of options open to us decreases, and we sometimes arrive at the fifth decision point only to discover that the data cannot be analysed statistically because there is no known test capable of handling data of the kind we have.

The problems of statistical analysis must to some extent determine the choice of sampling method since data which incorporate sampling biasses of unknown proportions cannot be rescued by any amount of statistical wizardry. In point of fact, the five decision points are not independent of each other, for the options which are open to us at point 5 may restrict the kinds of decisions we can make at points 3 and 2, and vice versa. Suppose we wished to investigate the relationship between dominance and grooming relationships. To do so we have to be able to compare two behavioural parameters, and before we can even decide what we are going to record, let alone how to quantify it, we are confronted with the spectre of a stage 5 decision. The availability of suitable statistical tests may limit the kinds of sampling strategies we can use, and these in turn may limit the behaviours sampled. Normally, of course, it is logical to proceed in the reverse order and make the necessary changes to previous decisions as we come up against insurmountable problems. However, the point serves to illustrate the fact that the constraints may come not only from above, but also from below.

Our main concern in this paper has been with stage 2 decisions, namely choosing the appropriate behavioural parameters to quantify in order to answer the research question posed. The critical problem at this point lies in deciding which particular behaviours are relevant to the question in hand. In many cases, the inherent difficulties are not at all obvious. Often it may seem that the topic we wish to investigate, for instance grooming relationships, leaves little room for doubt concerning the behaviours which ought to be quantified. Grooming, we suppose, is simply grooming, but there are at least three quite distinct problems here.

Firstly, we have to decide precisely what should be included in the activity of grooming, *i.e.* we have to arrive at an operational definition of grooming. Preferably, our definition should be related to those used by other authors, since the use of different definitions for the same term (especially when unstated) may lead to confusion in the literature.

The second problem concerns external validity. Are the behaviours we choose to quantify relevant to the question posed? In cases where the research question is very precisely defined there may be little problem here. However, where the research question is of a rather vague nature, the investigator runs the considerable risk of recording data which are inappropriate. As we have seen, there are at least three types of behaviour which might be considered relevant to a study of "social relationships", and which we choose to quantify will affect the results we get. Strictly speaking, each of these three parameters answers a different question. Analysis of grooming relationships answers the question "How do the animals distribute their grooming?", while an analysis of spatial relationships answers the question

"Do animals prefer to be spatially closer to some individuals than to others?". Neither necessarily tells us anything about the way in which the animals distribute their social contacts.

Thirdly, and finally, there are two quite distinct ways of quantifying any given behaviour, namely by frequency of occurrence (rate measures) and by duration (proportion of time measures). Which we choose may make a considerable difference both to the kinds of statistical analyses we can make and to the kinds of statements we can make about the animals' behaviour. We cannot necessarily assume, for instance, that grooming relationships are completely described by an analysis of the frequencies with which individuals groom each other. To the contrary, the lengths of time for which individuals are prepared to groom may be at least as interesting and probably just as significant.

In conclusion, then, four general points can be made relating to the choice of behavioural parameters to quantify. Firstly, the research question needs fine definition. If the problem is not clearly defined, we may find ourselves collecting data, which although admirable in themselves, do not answer the question we suppose ourselves to be asking. Secondly, the dictum that ethological studies should begin with a substantial period of naturalistic observation before any attempts are made to quantify aspects of a sociobiological system must be emphasised. Only in this way can the external validity of the research project be guaranteed. This applies both to the choice of a suitable problem to study (and, hence, an adequate definition of the research question) and to the choice of the correct behavioural parameters to quantify. Thirdly, the fact that there are often several behavioural parameters relevant to a given research topic suggests that it may often be useful to examine more than one facet of the problem. Finally, the interdependence of the five decisions makes it necessary to bear in mind, even when defining the research problem, aspects which might seem at first sight to be irrelevant at such an early stage. Not only must the availability of suitable statistical procedures be borne in mind, but so must the limitations of the various sampling strategies and the exigencies of the actual field situation (including the species being studied). Certain sampling strategies may be too difficult or too time-consuming to employ in a particular field situation, with the result that the kinds of data which can be quantified may be restricted. This, in turn, may limit the kinds of questions we can ask about the animals' behaviour.

SUMMARY

This paper has been concerned with examining the effects of choosing different behavioural parameters to investigate a particular research problem. Social relationships

among a set of animals were estimated in seven different ways. Three behavioural parameters (social contacts, grooming and spatial associations) were selected and a number of sampling strategies were used to quantify them. Although the measures yielded results which correlated significantly, there were marked numerical differences between them due to the fact that they measured different aspects of a complex biological system. The differences between frequency and durational measures were investigated and the accuracy of two common durational measures was determined. One-Zero sampling was found to provide a poor estimate of the proportion of time spent grooming, whereas instantaneous or point sampling always gave a reliable estimate. Finally, the relationships between the different stages of research design were discussed. Emphasis was laid on four aspects of choosing parameters to quantify, namely, the precise delineation of the research problem, the validity of the parameters in relation to the research question, the differences between the various measures and the requirements and limitations of the other key features of research design.

REFERENCES

Aldrich-Blake, F. P. G., Bunn, T. K., Dunbar, R. I. M. & Headley, P. M. (1971). Observations on baboons, *Papio anubis*, in an arid region in Ethiopia. — Folia primatol. 15, p. 1-35.

Altmann, J. (1974). Observational study of behaviour: sampling methods. — Behaviour 48, p. 1-41.

Altmann, S. A. (1968). Sociobiology of rhesus monkeys. III. The basic communication network. — Behaviour 32, p. 17-32.

— & Wagner, S. S. (1970). Estimating rates of behaviour from Hansen frequencies. — Primates 11, p. 181-183.

Bramblett, C. A. (1970). Coalitions among gelada baboons. — Primates 11, p. 327-334.

Carpenter, C. R. (1964). Social behaviour of non-human primates. — In: Naturalistic Behaviour of Non-Human Primates, C. R. Carpenter (Ed), p. 365-385. Pennsylvania State University Press, Pennsylvania.

Chalmers, N. R. (1968). The social behaviour of free living mangabeys in Uganda. — Folia primatol. 8, p. 263-281.

Deag, J. M. (1974). A study of the social behaviour and ecology of the wild Barbary Macaque, *Macaca sylvanus* L. — Ph.D. Thesis, University of Bristol.

Dunbar, R. I. M. & Dunbar E. P. (1975). Social dynamics of gelada baboons. Contrb. Primatol. Vol. 6. Basel: Karger.

— & Nathan, M. F. (1972). Social organization of the Guinea baboon, *Papio papio*. — Folia primatol. 17, p. 321-334.

Fienberg, S. E. (1972). On the use of Hansen frequencies for estimating rates of behaviour. — Primates 13, p. 323-325.

Hinde, R. A. (1973). On the design of checksheets. — Primates 14, p. 393-406.

Hutt, S. J. & Hutt, C. (1970). Direct observation and measurement of behaviour. — Thomas, Springfield.

Kummer, H. (1968). Social organization of Hamadryas Baboons. — Karger, Basel.

Leskes, A. & Acheson, N. H. (1971). Social organization of a free-ranging troop of black-and-white colobus monkeys (*Colobus abyssinicus*). — Proc. 3rd. int. Congr. Primatol. 3, p. 22-31. Karger, Basel.

Sade, D. S. (1965). Some aspects of parent-offspring and sibling relations in a group of rhesus monkeys, with a discussion of grooming. — Amer. J. phys. Anthrop. 23, p. 1-17.

— (1972). Sociometrics of *Macaca mulatta*. I. Linkages and cliques in grooming matrices. — Folia primatol. 18, p. 196-223.

Siegel, S. (1956). Nonparametric statistics for the behavioural sciences. — McGraw-Hill, New York.

SIMPSON, M. J. A. (1973). The social grooming of male chimpanzees. — In: Comparative Ecology and Behaviour of Primates, R. P. MICHAEL & J. H. CROOK (Eds), p. 411-505. Academic Press, London.

SPARKS, J. (1967). Allogrooming in primates: a review. — In: Primate Ethology, D. MORRIS Ed), p. 148-175. Weidenfield & Nicholson, London.

STRUHSAKER, T. T. (1967). Social structure among vervet monkeys (*Cercopithecus aethiops*). — Behaviour 29, p. 83-121.

RÉSUMÉ

Dans cette note, on examine les effets du choix des comportements divers pour l'investigation quantitative d'un problème socio-biologique particulier. Les rapports sociaux entre un groupe d'animaux ont été déterminés par sept méthodes différentes. Trois comportements (interactions sociales, épisodes du léchage et associations spatiales) ont été choisis et quelques méthodes ont été employées afin de les quantifier. Quoique les méthodes donnent des résultats qui s'accordent, on relève entre eux quelques différences numériques parce que chaque méthode considère l'aspect différent d'un système biologique compliqué. On examine les différences entre les mesures de fréquence et les mesures de durée, et on détermine la précision de deux mesures de la durée. La méthode „One Zero" donne une estimation imprécise de la proportion du temps que le sujet passe à lécher, mais la méthode instantanée ("point sampling") donne une estimation précise dans tous les cas. Enfin, on discute des relations entre les aspects divers de la méthode de recherches. On souligne quatre aspects du choix des paramètres à quantifier, à savoir la définition précise du problème des recherches, la validité des paramètres en relation avec la question posée, les différences entre les diverses mesures et les exigences et les limitations des autres caractéristiques des recherches entreprises.