Observational Study of Behavior: Sampling Methods

Author(s): Jeanne Altmann

Source: *Behaviour*, 1974, Vol. 49, No. 3/4 (1974), pp. 227-267

Published by: Brill

Stable URL: http://www.jstor.com/stable/4533591

# OBSERVATIONAL STUDY OF BEHAVIOR: SAMPLING METHODS

by

JEANNE ALTMANN [1)]

(Allee Laboratory of Animal Behavior, University of Chicago,
Chicago, Illinois, U.S.A.)

## CONTENTS

## I. INTRODUCTION

This is an observer's guide: in it I will present sampling methods for use in direct observation of spontaneous social behavior in groups of men or other animals. All observational sampling methods known to me will be described, and their uses and limitations indicated.

### A. SCOPE

In what follows I shall assume that the observer has a group of spontaneously interacting individuals to watch, that he has formulated one or more questions about social behavior, that he knows what behaviors he wishes to study, and that he has found suitable methods for recording such behaviors.

Thus, I will not be concerned here with the logistics of such research nor with the problems of choosing research questions and defining behavior categories. No attempt will be made to cover statistical aspects of experimental design, such as adequacy of sample size, factorial design, and so forth. Instead, the focus will be on a question that arises at an earlier stage in research planning, namely, how does the choice of sampling method restrict the kinds of behavior processes that can be studied? Or, conversely, given a particular behavioral research problem what sampling methods are suitable for it?

I shall assume that the research question has been sufficiently well-formulated that the investigator can identify the relevant sample space, *i.e.* the set of events that must be sampled in order to answer the question. Let me illustrate. Suppose that we are interested in studying aggression in a group of monkeys. We might begin by formulating a question such as this: Are males more aggressive than females? At this stage, the question is too ambiguous for a sampling method to be chosen. For example, we need to specify which behaviors will be classed as aggressive, and which will not. We need to make clear whether the question refers to all age classes or, say, just to adults. Even then, there are numerous reasonable interpretations of the question, such as:

(i)   On the average, do males spend more of the day involved in aggressive behavior than do females?

(ii)  Do males initiate aggressive bouts more often than do females?

(iii) Are the aggressive acts of males more serious, more intense, more potentially destructive?

(iv)  Do the behavioral acts of males include a higher proportion of aggressive acts than do those of females?

(v)   Is the response to an aggressive act more likely to be an aggressive act if the recipient is a male?

The choice of one or more of such formulations depends on an evaluation of their relevance to the original question. That evaluation depends, in turn, on numerous questions about the behavioral or biological significance of sex differences in aggression. Such evaluation lies outside the scope of this paper. But an unambiguous formulation of the research question is a prerequisite to the kinds of sampling decisions that will be discussed here: different formulations will usually involve sampling from different sample spaces.

The observer needs to know how to gather data that will answer such specific research questions. Should he repeatedly scan the group, recording

each individual's behavior at the time that it is seen? Should he watch each individual in turn, each for a fixed amount of time? If an individual continues in one observation period the behavior that he began in a previous period, should the behavior be recorded again? Should the observer record every occurrence of a behavior, or only whether it occurred at least once in the observation period? In this paper, I shall examine such sampling alternatives and provide guidelines for choosing among them in observational studies.

Sampling decisions are made whenever the student of social behavior cannot continuously observe and record all of the behavior of all of the members of a social group, and must therefore settle for a partial record. However, even in the most systematic of observational behavior studies, only partial descriptions of the sampling procedure are provided. Seldom has an author provided justification for his choice of sampling method.

We suspect that the investigator often chooses a sampling procedure without being aware that he is making a choice. Of course, he does not thereby escape the consequences of the choice: the data that result from any sampling method can only answer certain classes of questions. From the standpoint of the behavioral questions, a given question can be answered only by data obtained through certain kinds of sampling methods.

## B. MANIPULATIVE *VS* NON-MANIPULATIVE RESEARCH

The method of direct observation plays a curious and unique role in the behavioral sciences. It is at once the necessary link between laboratory research and "real-world" behavior, and the bane of our aspirations for more accurate, more objective information about behavior.

From time to time, one hears the claim that accurate studies of behavior can be made only in the laboratory, and that quantitative research on behavior is not practicable in the context of ongoing, real-life situations. Such a restriction on research would mean that the behavioral sciences would forever forsake any hope of knowing whether their most powerful theories have any relevance to the world of behavior outside the laboratory. Unless we develop methods for field research that are comparable in sensitivity to those of the laboratory, the behavioral sciences will become progressively more isolated from the very behavior that their theories are supposed to explain.

A primary function of research design is to maximize the validity of the conclusions (WEBB *et al.*, 1966), *i.e.*, to minimize the number of plausible alternative hypotheses that are consistent with the data. It is useful to distinguish between internal validity, which deals with statements about the sample, and external validity, which deals with interpretations and generalizations from the sample to other situations or populations. For example,

housing experimental animals under identical conditions would represent an attempt to increase internal validity, whereas the process by which these animals were chosen in the first place or the similarity of the housing to their natural habitat would affect the external validity.

Internal validity is an essential component of external validity: to the extent that we have not eliminated alternative explanations for the results within our sample, we cannot rule them out of any generalization or interpretation derived from the sample. However, some conclusions depend more heavily on the generality (external validity) of the results than do others: internal validity should not be purchased through complete loss of external validity.

Laboratory research on behavior has usually emphasized internal validity, but in such research we ignore at our peril the question of whether high internal validity has been gained by an inordinate sacrifice of generality and relevance. The choice of animals and the artificial world in which they are placed may distort the results, or the experimental task that is presented to the animals may be largely irrelevant to an understanding of how these animals solve their own problems.

In contrast, observational field studies of behavior tend to show the converse imbalance: low internal validity but, ostensibly, high external validity. A major source of these imbalances is this: external validity has been largely ignored in laboratory studies, and internal validity, in field research. The assumption apparently has been that internal validity requires manipulation, by the scientist, of subjects and behavior and that external validity depends on a naturalistic setting and the absence of such manipulations.

Attempts to correct this imbalance have recently been made by utilizing information from field studies to design laboratory experiments (*e.g.*, MASON & EPPLE, 1969) and by bringing some of the manipulations of a laboratory experiment into the natural field situation (*e.g.*, HALL, 1965; KUMMER, 1971; MENZEL, 1969). This approach is one way, but not the only way, of increasing the internal validity of field research; it will not be discussed in what follows.

The primary function of experimental controls is to reduce or eliminate alternate hypotheses, and it is this general function of methodology to which the field worker should look in his attempt to increase the internal validity of non-experimental field studies. Observational research may require the development of research tactics that are particularly suited to its needs. Of these needs, one of the most important is for sensitive, non-destructive methods of studying social processes (BARKER, 1963).

As SCHNEIRLA (1950) has pointed out, controls are not absent in field situations; instead, they are usually "observation-selective" rather than "manipulative". What, then, are the non-manipulative controls that are available? Can we use them in such a way that we increase the internal validity of observational studies of behavior, without losing external validity?

Sampling decisions offer the student of behavior a prime opportunity to increase internal validity through means which are non-manipulative and are therefore less likely to alter or destroy the social system that is being studied. Use of such controls — in particular, sampling decisions — in observational studies of social groups can increase the validity of comparisons both within and between studies, whether observational or experimental, field or laboratory.

## II. SAMPLING VARIABLES

Before turning to specific sampling techniques, let us consider the major variables that distinguish existing sampling methods and that are most crucial for sampling decisions. As noted previously it will be assumed that the observer has a well-formulated research question, and that he has at least a preliminary catalog of behaviors that are under study. It will also be assumed that certain preliminary sampling decisions have been made: the study locale and population have been selected.

### A. BEHAVIOR RECORD

Events *vs* states.

Behaviors may be regarded either as events or as states. Events are instantaneous; states have appreciable durations. Of course, in reality, the performance of any kind of behavior takes some amount of time, however brief. But if we consider behaviors at the moment of their onset, or at any other single defining instant, then we are recording events. We can, for example, record that an animal assumes a sitting posture, an act that occurs at an instant (and is therefore an event), or that the animal is seated (a state).

Our choice between regarding behaviors as states or as events depends upon the questions about behavior that we are attempting to answer. In particular, questions about frequencies of behavior, such as questions ii, iv, and v on p. 228, entail considering the behavior as events[2]). Once the

2) The term 'frequency' will be used in this paper to mean *number of occurrences,* in accord with convention in the statistical literature. It has a different meaning in some other contexts. Thus, 'gene frequency' is used by population geneticists to refer to a relative frequency or *proportion.* In the physical sciences, 'frequency' commonly refers

investigator has decided on a defining event, such as onset, for a particular behavior, that behavior is not scored in a sample session unless the defining event occurs during the session, even though the behavior is otherwise "in progress" during the session. On the other hand, any question involving the duration of a behavior, or the percent of time spent in some activity (*e.g.* question iii, p. 228) is a question about states. To answer questions involving duration one could time each occurrence directly, perhaps using a standard stopwatch. Alternatively, if an exclusive and exhaustive classification of states has been made, one could record transition times (*i.e.* onsets and terminations), thus preserving frequency and sequence information as well as that of durations and time spent in various activities. For information only on percent of time spent in a particular state, one could merely cumulate durations of the state of interest (*e.g.* by means of a cumulative stopwatch) and record the total sample time. The extent to which various sampling techniques are suited to answering each of these two basic types of questions will be discussed in the sections on individual techniques.

## Completeness of frequency record.

If events rather than states are scored, the sampling procedures may be divided into three classes with respect to the completeness with which frequencies of behaviors are recorded. In one class of procedures (see, *e.g.*, Sections V and VI), a complete or total frequency record is kept. By this it is meant that during a sample, all occurrences of the behavior of interest for some subset of group members are recorded. In a second type of sampling, partial frequencies are obtained; such records usually consist of an unknown percentage of the total occurrences, which is variable from sample to sample, and from individual to individual. *Ad Lib. Sampling* (Section III) usually results in such records. Finally, the observer may record, during each sample period, the fact that the behavior occurred at least once (scored as *one*) or did not occur (scored as *zero*). Thus, a score of "one" may represent one occurrence or a multitude of occurrences. Such sampling is discussed under the heading "One-zero Sampling" (Section VIII).

## Frequency and rate.

In the behavioral literature, many comparisons that are presented in terms of frequency differences are actually statements about rates. For example, we cannot justifiably claim that the dominant male copulated more often

_____

to number of occurrences per second, and thus to a *rate*. In some common English expressions (*e.g.* "I frequently sunbathe") it seems that the intended meaning is sometimes *percent of time,* and other times *rate, i.e.* number of occurrences per unit time.

than did the subordinate, until we know the amount of time that was devoted to sampling the behavior of each, or, at least, that the two amounts were equal. The question that is being answered is whether the first male copulated more frequently *per unit time* than did the other, *i.e.* we are comparing rates.

If one knew that observation times were equal for all individuals, then the frequencies themselves could be compared directly, ignoring the time base if rates were not of interest. However, particularly in field situations, individuals are seldom observed for exactly the same amount of time, and often for very different amounts of time. This may be the result of circumstances beyond the observer's control, or a direct result of the sampling procedure. Under these circumstances, rate comparisons are the obvious solution. Thus, implicitly (if observation times are equal) or explicitly, rates are often being used in behavior studies. Time is often an important variable.

There are, however, other statements or questions about behavior that are based on frequency *per se,* and not on rates. Most such statements are essentially statements about conditional probabilities, in which the condition is the occurrence of acts of a particular kind. For example, the statement, "Males hit more often than do females," is a statement about rates, whereas, "The aggressive acts of males more often involve hitting than do those of females," is not: the latter is equivalent to, "The probability of hitting in aggressive acts by males is higher than it is in aggressive acts by females." Here is another example: the statement, "Adult males threaten juveniles more often than they threaten infants," is based on a comparison of rates, whereas, "Infants run away, when threatened by adult males, more often than do juveniles," is not [3]).

Since some of the sampling methods that will be described in this paper can be used to estimate rates of interaction and others cannot, it is important that the investigator know in each case which type of question he is posing.

Content of record.

Behavioral records may contain records only of the occurrences of behaviors of interest, or they may in addition, include a variety of other data. A record is usually made of the date and time of each sample session onset and termination. These times determine the length of intervals between samples, the duration of the samples and the seasonal and diurnal distribution of the samples, all of which are important. Particularly common in studies of social behavior are records of: (1) the actors, (2) the receivers or object

---

3) Questions that involve true frequency comparisons may still require consideration of time base, in order to insure an unbiased sample of the conditional events.

individuals, (3) the sequence of events or states, without times of occurrence, (4) the time of occurrence, thus also including sequential information, (5) duration of behavioral state (see p. 231-232), without recording the actual time of onset and termination, or (6) onsets and terminations for some or all behaviors of interest. In addition, the record may include contextual data, such as habitat, weather, predominant group activity, distances to or identities of neighbors, or the size and membership of the subgroup in which behavior occurred. As we shall see, the choice among such characteristics of the data will, in turn, narrow the choice of appropriate sampling techniques.

## B. SCHEDULING SAMPLE SESSIONS

### Scheduling of session onsets.

A sample session may be scheduled to begin at a predetermined time. One possibility is that the sample onsets are chosen as a stratified random sample (*e.g.* with a fixed number of samples per hour, beginning at times chosen randomly within each hour); another is that they are scheduled at a regular time (say, once an hour, on the hour, for all daylight hours), or after a fixed time has elapsed since the termination of the previous sample. Alternatively, the sample sessions may be scheduled to begin, not at a particular time, but whenever a particular behavior occurs (*e.g.* whenever a particular pair of individuals interact, or whenever the animals enter a particular habitat). Finally, there may be no scheduling rule: observations may be made on an "*ad lib.*" basis. It is highly unlikely that the scheduling of such *ad lib.* samples will ever be independent of the behavior (see Section III). Furthermore, the nature and extent of the dependence is less likely to be known than in the behavior-dependent case just mentioned.

### Scheduling session terminations.

A sample session may be scheduled to terminate after a fixed time period, after occurrence of a fixed number of behaviors, after a particular class of behaviors or interactions has terminated, or it may continue so long as the animals are in view. Alternatively, there may be no termination rule. Once again, such *ad lib.* termination produces a sample with unknown and perhaps variable dependence on the behavior being sampled. For example, ALTMANN & ALTMANN (1970) suggested that a mid-day peak in their data on baboon social behavior might be due to the fact that the observers were more likely to take a mid-day break if nothing of special interest was occurring. As a result, those mid-day periods during which observations were made probably were a biased sample of mid-day periods, and behavior that was recorded then probably was a biased sample of mid-day behavior.

Number of individuals per session.

If all occurrences of behaviors of interest are recorded for a particular individual during an entire sample period, that individual will be referred to as a *focal individual* for that sample period. Or, there may be a focal sub-group, which can range in size from one individual to the entire group. Alternatively, there may be no focal individuals, either because only partial frequencies are being recorded or because attention is focused on first one individual, then another, the choice usually determined on an *ad lib.* basis throughout the sample session. Any form of *Ad Lib.* Sampling introduces the problem of unknown and probably variable biases mentioned previously (See also Section III).

Selecting focal individuals.

The choices among potential focal individuals or among focal sub-groups can be randomized utilizing a table of random numbers (*e.g.* with individuals picked at random from all individuals in the group), it may be a stratified random sample (*e.g.* with a predetermined number of focal individuals randomly chosen within each of a fixed number of age-sex classes), it may be regular (*e.g.* rotating according to a fixed schedule through all of the individuals or all the individuals of a class), or irregular, with the individual chosen on the basis of some behavioral criterion, for example, the first pair to interact (see *e.g.* Section VII, Sequence Sampling) or the closest readily-visible individual. Such behavior-determined selection of individuals may decrease the amount of sample time in which no behaviors of interest occur, but it will also introduce dependence between the samples of behavior and of participants. The choice of selection criteria can best be determined by the demands of the particular research question.

### III. *AD LIBITUM* SAMPLING

In field studies of behavior, perhaps the most common form of behavior record consists of what I shall call "typical field notes", or "*Ad Lib.* Sampling." Of course, the same type of record can be obtained in the laboratory by means of non-systematic sampling or informal observations. Such records are the result of unconscious sampling decisions, often with the observer recording "as much as he can" or whatever is most readily observed of the social behavior of a group in which behaviors, individuals and often the times for behavior sessions are chosen on an *ad libitum* basis. Presumably this is the technique used in most observational studies in which no mention is made of the sampling method.

Two kinds of assumptions often seem to be implicit in attempts to utilize

data obtained from *Ad Lib*. Sampling for the purposes of quantitative anylysis of behavior. The frequency of two classes of behavior may be compared, as in: "Female A grooms more often that she threatens," or "Rhesus monkeys groom more often than they fight." The assumption made here is that the chance that a behavior would be recorded does not depend upon the class of behavior — that grooming is no more likely to attract attention than fighting, or at least that the magnitude of such a difference is negligible, relative to the actual differences in the frequencies with which the behaviors occur. Second, comparisons may be made across age-sex classes, as in: "Adult females groom more frequently than adult males." Here the assumption is that, for this class of behavior, the likelihood that a behavior would be observed and recorded does not depend on the age-sex class of the individual involved. Of course, a statement may involve simultaneously a comparison across behaviors and across age-sex classes, as in "Males fight more than females groom." Such conclusions involve simultaneously making both of the above assumptions about lack of bias, neither of which can be justified with this sampling technique.

These uses of *ad lib*. data are not, of course, the only ones. They have been pursued here as examples because they are common, particularly in first stages of quantification in field studies. The same line of argument can be used in making explicit the assumptions underlying any use of data. If data are to be used to answer a particular question, the observer should ask, What are the assumptions underlying such use and for which sampling techniques are such assumptions reasonable? Asking whether the assumptions are reasonable is just another way of looking at the question of whether alternative hypotheses have been ruled out. If, for example, *Ad Lib*. Sampling of behavior yielded a higher frequency of aggression by males than females, the results might be explained by the greater conspicuousness of male aggression, unless it is reasonable to assume that the sampling was independent of the sex of the actors.

With *Ad Lib*. Sampling, it is rarely possible to determine which differences in data are due to true differences between individuals, age-sex classes, or behaviors, and which due merely to biases in sampling. When comparing the results of one such study with those of another, we cannot tell which differences were due to differences in what could be seen, which to differences in what was selected for recording, and which to actual differences in the populations.

In any field study some data probably will consist of such records, which may be of considerable use as illustrative material and because .of their heuristic value in searching for ideas and in planning systematic sampling

of behavior. Often, too, rare but significant events are recorded during such nonsystematic sampling periods. But studies which consist only, or even primarily, of such records leave open too many alternative hypotheses that might account for the data. Without some form of systematic sampling procedure, there appears to be no way to avoid the bias that results when the observer's attention is attracted by certain types of behavior or certain classes of individuals.

If we could assume that the biases in *Ad Lib.* Sampling were of constant direction and magnitude over time, within and between studies, we might be able to compare such data despite these biases. Unfortunately, such an assumption will seldom be justifiable. This becomes particularly significant when one examines, as we shall in what follows, the unsuccessful attempts to correct for bias in such sampling in order to utilize the results.

Several authors (*e.g.*, ALDRICH-BLAKE, 1970; CHALMERS, 1968a; SADE, 1966) have suggested that a major bias in *Ad Lib.* Sampling of individuals results from the fact that some members of a group are more readily observed than others, and that this bias results directly from differences in the proportion of time that each individual is visible, rather than from individual differences in, say, size or activity level. CHALMERS (1968a) and SADE (1966) have attempted to provide a measure of such differential visibility.

In CHALMERS' (1968b) study of mangabeys in Uganda, he, too, attempted to measure differential visibility for various age-sex classes. He writes:

"Censuses were taken at half hourly intervals on the monkeys visible. These noted, among other things, the number of monkeys of each age/ sex class in sight."

CHALMERS compared frequency of appearance in such "censuses" (probably Instantaneous or Scan Samples, see Section IX) with that expected on the basis of group composition (as determined by an independent sample) and found significant deviation. CHALMERS next calculated the frequencies of various social behaviors that would be expected if the "census" frequencies adequately represented the differential observability, if differential observability were the only source of bias in sampling individuals and if animals of any one age-sex class were as likely as those of any other to engage in such behavior. He compared the expected with the observed frequencies and suggested that deviation from the "census" distribution provides evidence that members of the age-sex classes differed in the frequency with which they engaged in such behavior.

In his studies of the rhesus monkeys of Cayo Santiago, SADE (1966) measured differential observability as follows:

> "During each hour of observation a two minute period was chosen at
> random during which I noted down each individual member of Group F
> that I could find and identify."

These samples were taken by the observer as he walked through the group.
SADE refers to such records as "time samples" (see footnote p. 252). He
considers the results of such "time samples" to be a measure of the likelihood
that a record will be obtained if an animal did some behavior of interest
during observations of behavior, and that these "time sample" results can
therefore be used to correct for individual sampling bias.

In the only published attempt to utilize such "time sample" data, HAUS-
FATER (1971, 1972 and personal communication) proceeded as follows. For
each animal, he took the number of "time samples" in which it was seen
and divided that number into the smallest non-zero number of this kind in
the population. This gives a putative correction factor: when the number of
encounters in which the individual was observed to participate is multiplied
by this factor, a new number is obtained that is presumed to be the individual's
corrected number of encounters relative to other individuals (except those
individuals that were never observed during time samples).

Both of these methods are attempts to deal with and correct biased sampling
of individuals. These authors have singled out one possible component of
such bias: the fact that some individuals are available for observation more
of the time than are others.

The implicit assumption is that there exists a positive number, $K$ (an
"observability constant"), such that for each animal, $i$, $K$ times $p(O_i)$, the
probability that $i$ will appear during an observability sample, is equal to the
probability that a record will be made, during a behavior sample of $i$'s
participation in a behavior under study. It then follows that if $n_i/N$, the
proportion of observability samples in which $i$ appeared, is used as an
estimate of $p(O_i)$ for each individual that appeared in at least one obser-
vability sample; then multiplying $N/Kn_i$ times $x_i$, the observed number of
participations by $i$ during behavior sampling will yield an estimate of the true
number of participations by $i$. (Multiplying all of these values by any non-
zero constant will yield *relative* rather than true participations. HAUSFATER
(1971, 1972) did this by multiplying them by $n_m K/N$ where $n_m$ is the
frequency of appearance in observability samples for the animal that appeared
least often, but at least once.)

But what are the grounds for this basic assumption? In order to justify
the use of observability samples, one would require either direct testing of
this basic assumption or testing of other assumptions from which it could

be derived. No such direct or indirect test has been published. Consider the following line of argument, which may represent the rationale that was used:

(i) The observability sample is assumed to provide an accurate measure of relative observability in behavior samples with which it is used, *i.e.*, the proportion of time an animal is visible during behavior sampling is assumed to be adequately estimated by the proportion of Observability Samples in which he appears. Procedural differences in these two types of sampling can result in a difference between the relative amount of time that an individual is visible in each and, therefore, in failure of this assumption. For example, such differences in observability will be present to the extent that the observer scans the group differently in the two kinds of samples.

(ii) Even if the observability samples provide an accurate measure of the proportion of time that each individual was visible during behavior samples, we still need to assume that there is a relation between such observability and the probability that an individual's behavior would be recorded. Observability samples provide an accurate correction only to the extent that the probability of a behavior being recorded if any individual performs that behavior is directly proportional to the percent of time that the individual is visible.

There are at least three sources of failure to obtain such a consistently proportional relationship in *Ad Lib*. Sampling. (a) Individual or class-specific differences in observability may vary greatly from one kind of behavior to another. For example, a subordinate male but not a dominant one might tend to mount in concealment, and thus would be obscured at such times, not just from other members of his group, but from the observer as well, despite the fact that both males may be visible to the observer the same percent of the day. (b) Some forms of behavior may affect the observability of any individual that participates in them. For example, extensive chases may bring both participants into view, thus making it irrelevant whether one of the participants is seen in, say, twice as many observability samples as is the other. (c) Any time that more is visible than can be recorded, sampling decisions remain and, in the absence of systematic sampling, the observer's preferences will come into play. Thus, if some individuals or behaviors (or a combination of the two) are rarely seen, the observer may unconsciously "compensate" by paying more attention and recording more when they are seen. This might, for example, lead to an overestimate of the frequency, or relative frequency, of a rarely seen form of behavior, particularly if "correction" factors were applied to the data.

Under any of these circumstances, application of a correction factor may

not merely fail to provide a complete correction: it may increase the existing bias. The result would be a poorer estimate of what actually occurred than the original data would have provided.

In summary, then, the application of observability correction involves a number of assumptions. Justifying the use of these or similar corrections would require evidence that these assumptions were reasonable under the sampling and behavioral conditions, or at least, that the use of the observability corrections would result in a reduction in bias over the amount in the *Ad Lib.* Sampling alone. One obvious source of such evidence for the validity of these corrections would be a comparison of their results with the results of a sampling method that is unbiased with respect to individuals. But then, the corrections will usually be superfluous. The more productive approach is to look for sampling techniques that are unbiased with respect to the main variables of interest. In what follows I shall consider the extent to which other existing techniques enable the observer to avoid various biases.

## IV. SOCIOMETRIC MATRIX COMPLETION

In some studies, *Ad Lib.* Sampling has been supplemented by making additional observations on particular pairs of individuals. This has been accomplished by spending more time with these individuals, or, *e.g.,* by experimentally provoking a fight by means of competition in pairs of individuals for whom the original sample size was considered inadequate. The results of such sampling are usually published in the form of a "sociometric matrix," that is, a contingency table in which actors (*e.g.,* aggressors, or winners) are represented by the rows, and recipients (*e.g.,* losers of fights) by the columns, and in which the cell entries indicate the frequencies of the corresponding (dyadic) interactions. In these studies the object has been to establish, for each pair, the direction and degree of one-sidedness of some relationship, such as groomer-groomee or winner-loser of fights (see *e.g.,* ALEXANDER & BOWERS, 1969; SADE, 1966; MISSAKIAN, 1972, in which this technique apparently was used). In such studies, then, a sample usually is considered to be inadequate if the data in a pair of cells are small in number, or large in number but nearly equal.

When data for such a sociometric matrix are obtained by this sampling procedure, the result is not a contingency table in the usual sense, but simply a compact way of tabulating data. No biological interpretation can be given to the row or column totals. Each cell frequency reflects both the effects of the animals' choices among partners in dyadic interactions and the effects of attempts by the observer to boost the frequencies of certain cells. Consequently, one cannot directly compare each cell with every other cell: they

do not repersent the results of unbiased sampling of dyadic interactions. Another consequence is that the row totals are probably a biased sample of the distribution of acts by the members of the group; similarly, the column totals are probably a biased sample of acts received by the members of the group. Any row or column will contain such biases if it includes cells some of whose data result from the supplementary observations.

Certain kinds of questions cannot in general be answered by such data, *e.g.* for any two individuals A and B, "Does individual A do more grooming than individual B?," or even, "If A grooms, is he more likely to groom B than C?" The former would have to be answered through unbiased samples of grooming bouts (or, at least, of groomers) in the group as a whole, the latter, by at least an unbiased sample of A's grooming bouts.

However, if, for each pair of individuals, the observer can assume that the data represent an unbiased sample of their relations in their paired encounters (*e.g.* grooming sessions between A & B), that the outcome (A grooms B, or B grooms A) of any trial (grooming event) is independent of the outcome of any other, and that the probability of each outcome remains constant from trial to trial, then each cell of the matrix can be compared with the corresponding cell of the transposed matrix, treating the two cells as the components of a binomial distribution. For example, the frequency with which any individual, A, groomed any other individual, B, can be compared with the frequency with which B groomed A. We could also ask whether A is more likely to be the groomer in his grooming interactions with B than in those with C. We could look at the "linearity" (transivity) of grooming; *e.g.* if, for any three members of the group, A, B, and C, among whom A usually grooms B (rather than vice versa), and B usually grooms C, is it true that A usually grooms C more often than C grooms A? As a note of caution, we observe that the latter two questions will usually involve comparisons of binomial probability estimates with different sample sizes and hence different confidence intervals.

It would be preferable, then, to present such data in the form of a table that brings together the data for each pair of individuals, thereby facilitating binomial testing or other comparisons and avoiding the temptation to treat the data as if they constituted entries in a conventional contingency table or matrix.

In summary, the technique of Matrix Completion of *Ad Lib.* Samples is particularly suited to studies in which the basic problem of interest is the direction and degree of one-sidedness in the relations of each pair of individuals, but is ill-suited to answering many other types of questions about behavior. If the observer's main interest is in such asymmetry problems,

if he feels that the binomial assumptions are satisfied, *and* if he feels he can obtain much larger sample sizes with this technique, then it might be the technique of choice. One alternative, Focal Animal Sampling (Section V), would provide relatively unbiased data both on degree of asymmetry and on many other aspects of behavior as well.

Perhaps the best solution would be to begin with Focal Animal Sampling (rather than *Ad Lib.* Sampling) and then to supplement these data as needed to insure an adequate sample for each pair. This supplementary sampling might consist of additional Focal Sampling of particular individuals. If doing so required that unacceptably large amounts of time be devoted to those individuals, then the observer might work on other aspects of the study until individuals of a pair in question moved near each other, at which time sampling on them would begin. The data from the Focal Animal Sampling would then be available for other kinds of analysis for which its relative lack of bias would be advantageous.

## V. FOCAL-ANIMAL SAMPLING

I use the term Focal-Animal Sampling to refer to any sampling method in which (i) all occurrences of specified (inter)actions of an individual, or specified group of individuals, are recorded during each sample period, and (ii) a record is made of the length of each sample period and, for each focal individual, the amount of time during the sample that it is actually in view. Once chosen, a focal individual is followed to whatever extent possible during each of his sample periods.

This kind of sampling has been used in field situations, by BEER (1961-63) in studies of gulls, by WOOTTON (1972) in studies of stickleback fish, by STRUHSAKER in studies of elk (1967) and vervet monkeys (1971), by FISLER while observing Cayo Santiago rhesus (1967), by RICHARD (1970) in a study of howler and spider monkeys, by SAAYMAN (1972) in studies of chacma baboons, and by ALTMANN & ALTMANN (in preparation) in studies of social behavior of Amboseli baboons. The work of DOYLE *et al.* (1969) on lesser bushbabies, of PLOOG and his colleagues (PLOOG *et al.*, 1963; PLOOG, 1967; HOPF, 1972) on squirrel monkeys, of ROSENBLUM & KAUF-MAN (1967) on pigtail and bonnet macaques and of SCRUTTON & HERBERT (1970) on talapoin monkeys, provide examples of its use in studies of caged colonies. Examples from research on human behavior include the work of SWAN (1938), WASHBURN (1932), Mapheus SMITH (1931, 1933), CHAPPLE and his colleagues (1963), BLURTON-JONES (1967, 1972), and McGREW (1972).

In a Focal-Animal Sample, the sampling of non-social behavior is rela-

tively straightforward. I shall discuss below some of the problems of sampling social behavior. Most such behavior is directed ("addressed"); I shall distinguish between the actor or sender, and the object or receiver of each social act.

Under some conditions and at least for some behaviors, one may reasonably assume that a complete record is obtained not only of the focal animal's actions, but also of behaviors directed to him by others. Then a Focal-Animal Sample on animal $i$ provides a record of all acts in which $i$ is either the actor *or receiver*. This means that both during animals $i$'s focal samples and animal $j$'s focal samples we are recording all interactions between $i$ and $j$: either sample or both together would provide the necessary data for estimating their rate of interaction (see Frequencies and Rates, p. 232, 233).

Under other circumstances, it may be possible to record all acts by the focal individual, but not all those directed toward him by others (*e.g.* silent threats). For those behaviors for which records are incomplete, it may still be reasonable to assume that the sample distribution of, say, senders is unbiased. However, to estimate the rate of interaction between the two animals, $i$ and $j$, we would then need to use $i$'s sample for those acts directed from $i$ to $j$, and $j$'s sample for the rate at which acts are received by $i$ from $j$ (see Frequencies and Rates, p. 232, 233).

Focal subgroups.

Although Focal-Animal Sampling as defined above does not exclude the possible use of a focal (sub)group of several animals, such sampling will usually be practicable only when it is possible to keep every member of the focal subgroup under continuous observation during the sample period. The reason for this is that the sample space in a focal sample consists of those dyadic interactions in which at least one participant is a member of the focal group. If only one of two focal animals were visible for, say, 5 minutes, that period can still be used in estimating the interaction rate between the two of them but if both were out of sight during some time, no interaction between the two would have been available for observation during that time. Consequently, focal group sampling requires that the concurrent observation time be known for every pair of focal individuals. (Similarly, such time records would be necessary for every triple of focal individuals if triadic interactions were also under study.) Under most circumstances, the only condition under which such a record can be obtained is that in which all the individuals in the sample group are continuously visible throughout the sample period.

Beyond the problem of time records, there is a further reason for having

only one focal individual per sample period. One of the great strengths of Focal-Animal Sampling is that in order to stay with a focal individual, the observer follows him and obtains observations on him in situations in which he would not ordinarily be under observation. This advantage would be lost by having multiple focal animals: if one of three focal individuals moved out of sight, pursuing that one would usually mean losing the others.

Thus, if one is working with observational conditions that are less than perfect, Focal-Animal Sampling should be done on just one focal individual at a time, or at most a pair (*e.g.* mother and young infant). The following discussion of the Focal-Animal Sampling method will assume that there is only one focal individual in each sample; however, all of the principles that are described also apply *mutatis mutandis* to Focal-Animal Sampling on subgroups and on whole groups.

Time records.

For some research problems, time may not be an important variable (*e.g.* in a study of the response to a particular behavior, or of the order of behaviors in a sequence of interactions). For such a study one might want to use Focal-Animal Sampling in which no record is kept of the length of the sample period or of the time that the focal individual is in view. However, many questions about behavior are known to involve a time base (See Frequencies and Rates, p. 232, 233). For others, we may not know ahead of time whether time is a variable that can be ignored. Thus, in an exploratory study, an observer will do well to record time information, even if internal behavior-conditional aspects of the activities are of primary interest, or the session onset and termination rules are behavioral ones (see p. 234).

The simplest method for obtaining time records in Focal-Animal Sampling is to sample for a predetermined amount of time, keeping records of the amount of time during each sample session that the focal individual is visible and being sampled ("time in"), or else of the amount of time that it is out of sight ("time out"). A cumulative stopwatch is useful for this purpose. Alternatively, the end of the sample period can be determined by other stopping rules, such as after a predetermined amount of "time in". Or sampling might be terminated according to some behavior-dependent stopping rule, such as after the behavior under study has occurred a predetermined number of times. Such a technique could be used to guarantee an adequate sample size (see p. 234). However, the observer should bear in mind the need for a stopping rule that is independent of the behavior parameters that will be investigated (p. 234). Independence might be obtainable with a behavior-dependent stopping rule if few and explicit research questions have been

formulated ahead of time and none will be added later. Otherwise, sampling for a predetermined amount of time will usually be the method of choice.

The choice of sample session length will depend upon several considerations. An upper limit on the length of the sample sessions will be set by observer fatigue. If sample sessions are too long, it becomes increasingly difficult to keep one's eyes and attention fixed on a single individual, and the accuracy of the records is affected. Of course, fatigue depends in part on one's familiarity with the species and its repertoire. Much greater mental effort is required to encode unfamiliar behavior. Fatigue will be affected by the number of behavior categories to be recorded, the rapidity and subtlety of these behaviors, and the amount of contextual and sequential information to be gathered.

If the durations of behavior are of interest, then the sessions should be long enough to obtain an adequate estimate of the distribution of durations. The differences between many common statistical distributions is revealed in the tails of the distributions (COCHRAN, 1954). Similarly, if sequential constraints are under investigation, the sample period should be long enough to include an adequate sample of the longest sequences that are of interest. If only frequencies are of interest, then the length of each session is theoretically immaterial. Of course, the total "time in" for all samples which are to be pooled must be long enough to provide adequate estimates for the least frequently occurring behavior under study.

Scheduling focal individuals.

Depending on the nature of the research problem, there are various possibilities for the focal individuals that will be covered in a study: all members of a group, all members of certain age-sex classes or some other subset (say, females with neonates), some members of particular subsets, and so on. Assumptions about variability within and between individuals and classes will affect scheduling decisions — whether, for example, one samples five males for twenty samples each, twenty males for five samples each, or even one hundred males for one sample each. Likewise, if the observer does not wish to assume that diurnal variability in behavior is negligible, focal individuals might be sampled at the same time of day or each one at several periods (say, once an hour) during the day. Otherwise, daily sample periods could be assigned to individuals without regard to this variable. Thus, the assignment of individuals to sample periods and the scheduling of sample periods will depend on both the questions being asked and the assumptions that one is willing to make (see also p. 235).

In field research, it is not always possible to recognize individuals. Under

those conditions, it is not feasible to make an unbiased selection of focal individual, to take individual variability into account, or to study such variability. Nevertheless, a kind of Focal-Animal Sampling can be carried out, choosing at random among visible individuals, utilizing a table of random numbers, then continuing to sample that individual so long as it is possible to keep track of him. Such a procedure might be preferable to resorting to other, even more biased sampling methods.

Behavior record.

When using Focal-Animal Sampling in studies of the social behavior of primates or other highly social animals, so much data may be obtained that one pushes the limit of the observer's ability to process information. I have already discussed the problem of observer fatigue. In addition, we have found, with the baboons and macaques that we have observed, that it is not possible simultaneously to record all social behaviors, their durations, the sender and receiver, the distance relations of the participants, their neighbor relationships, and the temporal pattern of the behaviors even for one focal individual per sample. Data on temporal patterning is particularly difficult to obtain: in one study (ALTMANN & ALTMANN, in prep.) we found that even with two observers, one 15-minute sample per hour was near the upper limit of our capacity when obtaining an accurate record, with some 5 dozen social behavior categories, of who did what to whom and in what order, as well as keeping track of most nonsocial behavior, durations, and time-out periods. In a more recent study in which we utilized 40-minute Focal-Animal Samples, we were able to collect extensive data on neighbor relationships and on social interactions (including the behavior, and the age-sex class of the social partner) of wild baboons, but without obtaining complete sequential records or much information about the durations of behaviors. In that study we took a ten-minute rest after such a sample and then reversed the roles of primary and secondary observers before taking a second 40-minute sample.

Montgomery SLATKIN (personal communication) has been able to obtain data on the durations of behavioral states in baboons, as part of his field study of baboon time budgets. He utilized the classification of the activity states of the individual into 5 exclusive and exhaustive categories. Then, during Focal-Animal Sampling, he recorded all transition times and the behavior state following each transition. Such records required two observers. The primary observer kept his eyes on the focal individual, punched a stop-watch at every transition, and dictated the behavior. The assistant drove the

vehicle and recorded the transition times and activities. (A stopwatch with an extra "marker" hand was invaluable for this study.)

In another study, Thomas STRUHSAKER was able to obtain data on the duration of behavioral states in focal mother-neonate pairs of vervet monkeys (STRUHSAKER, 1971).

In summary, with appropriate choice of focal individuals, sample periods and behavior records, Focal-Animal Sampling will usually be the technique of choice. It can provide relatively unbiased data relevant to a wide variety of questions about spontaneous social behavior in groups. Since observation is usually made on one animal per session, to the exclusion at those times of detailed information about other (inter)actions in the group, this technique is least suited to answering questions about behavioral synchrony. For most such questions, the methods discussed in Sections VI and IX would be more appropriate, but for studying behavioral synchrony among neighbors, Focal-Animal Sampling might be the method of choice.

## VI. SAMPLING ALL OCCURRENCES OF SOME BEHAVIORS

Under some conditions, it is possible to record all occurrences of certain classes of behaviors in all members of the group during each observation period. Such samples have been obtained by ROWELL (1967, 1968) in her studies of a caged social group of baboons, by CRAIG et al. (1969) in studies of agonistic behavior among birds in a field situation, and by LINDBERG (1971), who obtained data on the frequency of all agonistic vocalizations for 20 hours, in a field study of rhesus monkeys. Such records are generally possible only if (i) observational conditions are excellent, (ii) the behaviors are sufficiently "attention-attracting" that all cases will be observed, and (iii) the behavioral events never occur too frequently to record. For example, in our studies of baboons in Kenya, (ALTMANN & ALTMANN, in prep.) we were able to keep such frequency records for agonistic encounters that involved a vocalization and for sexual mounting between adults. Even for these two categories, we did not always obtain complete records of the individuals that were involved and certainly not of the complete sequences of behavior, since our notes on many of the occurrences began with the actual agonistic vocalization or when one individual was seen mounted on another.

For behaviors that can be sampled in this way, what kind of information can such sampling provide?

(i) With a wise choice of sample periods, it can provide accurate information about the rate of occurrence (and temporal changes in the rate) of such behavior in the group as a whole. If all participants can be identified at each occurrence of the behavior under study, this sampling technique is equi-

valent to Focal-Animal Sampling (Section V) on the whole group with respect to this particular behavior, and provides data of the kind discussed in that section. When not all identifications are possible, the data that do include identification will be an unbiased sample of the distributions of those behaviors among individuals (or classes) if there were no differential identifiability, *i.e.* if identifiability were random with respect to individuals (or independent of their age-sex classes). By the same token, they will be an unbiased sample of the outcomes of dyadic interactions, and could therefore be used to answer questions, of the sort described in Section IV, about the degree of one-sidedness of relations.

(ii) This is not the technique of choice for many kinds of sequential analysis. However, if the sequential information that is desired is the sequelae of some behavior that can be sampled in this way, *e.g.* the response to vocal threats, one could start each sample with an observation on such behavior, then record what happens next.

(iii) This sampling technique is appropriate for studies of behavioral synchrony if the observational and recording conditions are such that occurrences of the behavior can be recorded even if they are simultaneous. Actual time of occurrence, rather than just the frequency within an interval or the number of simultaneous occurrences, would, of course, provide more fine-grained information as to the temporal distribution of the behavior. This information may be of interest in itself, or it might be needed in order to test certain assumptions that are made when using other sampling techniques.

## VII. SEQUENCE SAMPLING

In Sequence Sampling, the focus of observation is an interaction sequence, rather than any particular individual(s). A sample period begins when an interaction begins. During the sample, all behaviors under study are recorded, in order of occurrence. The sample continues until the interaction sequence terminates or is interrupted, and the next sample begins with the onset of another sequence of interactions.

Sequence Sampling has been used in studies of social behavior in crabs (HAZLETT & BOSSERT, 1965) and monkeys (S. ALTMANN, 1965). In both studies, sequential dependencies in communicative interactions were of primary interest. HAZLETT & BOSSERT (1965: 359) write:

"To carry out the observations reported here, groups of 25 to 100 individuals were placed in an observation aquarium, and after 15 minutes, observations were started. Whenever the movements of one or two of the crabs were such that it appeared they would subsequently come

into social contact (HAZLETT, in press), recording was started. When one crab deviated from its path before the animals came into social contact, *i.e.* there was no observable interaction between the two crabs, the recording was stopped and discarded in later calculations. The movements and displays of the two interacting individuals were recorded until one or both crabs moved away from one another, either by some form of retreat, climbing over the other crab or by moving past one another. If the interacting pair was interrupted by a third individual, recording was stopped and the results discarded in later calculations."

Presumably, the first dyadic interaction to occur after observations began was the one that was recorded first, the next interaction that began after the first one terminated was recorded second, and so forth. Note that the Sequence Sampling of Hazlett and Bossert is essentially Focal-Animal Sampling (in which both members of the pair may be considered focal individuals) that differs from conventional Focal-Animal Sampling in that behavior-contingent rules were used for starting and stopping a sample. However, because of the effects of these rules on the records, the results will not be equivalent to Focal-Animal Sampling with, say time-contingent starting and stopping rules.

ALTMANN (1965) "tried to sample at random from among the monkeys," although "no systematic randomizing technique was used." If a selected individual was not interacting, ALTMANN chose another individual, continuing until an interacting individual was located. The sequence of interactions was then recorded, continuing "until the interaction process terminated, or until it was no longer possible for me to see everything that was going on." In that study, an interaction process was not judged to have terminated solely because the initial individual (or any other individual) left the interaction group, and thus this is not Focal-Animal Sampling, as we use that term (p. 242). For example, a play group might persist for some time, despite the fact that various individuals entered or left the play group during that period. The sequence of interactions within such an interacting group was recorded so long as it was, in some sense, unbroken.

Initially it seemed to me that this method was biased toward events that occur in sequences which, as a class, take up proportionately more time. Consequently, it would be biased toward acts, individuals, or sequential constraints that are different in such sequences. Thus, if two individuals spend the same amount of time interacting, but one is involved only in long interaction sequences and the other only in ones whose durations are half as long (so that there are twice as many of the latter) it appeared that the

observer would be just as likely to choose an interaction of the one as of the other, but he would then spend twice as long with (and record more acts of) the one involved in longer sequences. Similarly, if some behavior is more (or less) likely to occur in those longer sequences than in the short ones, it seemed that this sampling method would not provide good estimates of relative frequencies of behavior patterns. In this example, such sampling would therefore be biased toward the characteristics of events in long sequences.

Crucial in this line of argument are the probabilities of choosing sequences of various lengths. If the choice of the first (or any other) sequence to be sampled is made at random from among ongoing sequences, then the probabilities will in fact be equal to the relative amount of time taken up by the set of all sequences of that length (duration), as indicated in the example above. If, however, the observer always begins sampling at the onset of a sequence and chooses the next sequence to sample at random among sequence onsets or in any other way that samples sequences of each length in proportion to their *frequency of occurrence,* the resulting data will be unbiased with respect to sequence length: the total time spent with sequences of, say, duration $d_i$, will be proportional to $d_i$ times $f_i$, where $f_i$ is the frequency of sequences of length $d_i$. Then the time spent with sequences of different lengths, not the probability of choosing such sequences, will be in proportion to the total time taken up by sequences of that length.

Done this way, Sequence Sampling would still present several problems. It requires a method for choosing sequences that satisfies the above-mentioned criterion and a way of identifying the beginning and end of each sequence. If the observer always chooses the next available sequence onset, or one that occurs in the vicinity of the last one, the sampled sequences may not be independent of each other. Yet, it is not obvious how one might pick sequences at random. In addition, getting a record of a sequence from its beginning places heavy reliance on the ability of the observer to anticipate those circumstances under which interactions are likely to occur.

ALTMANN's method in particular illustrates both advantages and disadvantages of Sequence Sampling. The sampling procedure that was used enables the observer to stay with and record social interactions, the persistence of which does not require the continued participation of any one individual. For example, monkey A aggresses against monkey B, who redirects the aggression to C, who then enlists the collaboration of D. The method thereby provides information about the sequential structure of social interactions that is not provided by Focal-Animal Sampling or any other sampling procedure described here. Another advantage is that large

samples of social interactions can be obtained by this method: because the observer takes the next available interaction in the group, his time is seldom spent without available data.

However, specifying criteria for identifying the beginning and end of a sequence may be difficult. The sequence definition that S. ALTMANN used depends on the fact that the behavior of one individual may influence the behavior of another, which may, in turn, affect the next reaction, and so forth. The resulting chains of influence are Altmann's interaction sequences, and it is these chains that he followed in Sequence Sampling. Regarded in this way, interactions may have two properties that present sampling problems, branching and converging. If a sequence branches (e.g. if a play-group divides in two without a break in the interaction), which branch should the observer follow? Or, if two sequences should join into one (e.g. if an individual goes from one interacting group to another without a break) how can the conjoint influences be sampled? (HAZLETT and BOSSERT avoided both these problems by restricting their sampling to sequences involving interactions between just two individuals.)

In summary, the primary advantage of Sequence Sampling is that it enables the observer to obtain large samples of social behavior and to sample sequences of interaction that may persist regardless of the continued participation of any one individual. The primary disadvantages stem from problems in selecting sequences and identifying their beginning and end.

## VIII. ONE-ZERO SAMPLING

### A. BACKGROUND

During the 1920's a systematic sampling method was developed for studying spontaneous behavior in children, (OLSON, 1929), and was referred to as "time sampling." GOODENOUGH, one of the earliest advocates of this sampling method, defined "time sampling" as:

> "... the observation of everyday behavior of an individual or a group of individuals for definite short periods of time and the recording of the occurrence or non-occurrence of certain specified and objectively defined forms of behavior during each of these periods." (GOOD-ENOUGH, 1928, p. 23.)

The common features of the technique are the following. (i) In each sample period, occurrence or non-occurrence (rather than frequency) is scored. (ii) Interactions of just a single individual or pair of individuals are recorded in each sample period. (iii) *Occurrence*, for most users of this technique,

has meant "in process" at any time during the sample period, *i.e.* a sampling of states rather than events (see p. 231). (iv). The sample periods are usually short (*e.g.* 15 secs.), with about 20 sample periods in succession. Such batches of samples may then be repeated, perhaps twice a day over the period of the study [4]).

Observational studies of children in social groups were relatively common for several years, particularly during the 1930's (WRIGHT, 1960). In such research, "time sampling" (*i.e.* One-Zero Sampling) predominated. However, only one of these early studies (OLSON, 1929) utilized One-Zero Sampling for animal investigations.

During this early period of development, a number of workers investigated methodological questions. Those who criticized "time sampling," as well as proponents of the technique, focused on secondary questions, such as the appropriate length for the sample periods, adequate sample sizes, changes in the state of the system over time, observer agreement, choice of behavior categories, and so forth (cf. ARRINGTON, 1943; M. SMITH, 1931; OLSON & CUNNINGHAM, 1934). At no point, however, was the basic rationale for One-Zero Scores questioned.

In later years, observational studies became relatively less common in child behavior research (WRIGHT, 1960). Among studies that continued to utilize observational techniques, increased emphasis was placed on rating scales and on controlled, one- or two-person settings. Observers tended to use interpretive behavior categories, such as *seeks attention,* rather than relatively non-interpretive motor patterns, such as *hits.* A few workers turned to other observational sampling techniques (see e.g. Sections VII and IX), while others (*e.g.* BISHOP, 1951) continued to use One-Zero Sampling.

Observational studies, of both human and non-human behavior, have become increasingly popular in recent years. In the study of animal behavior, One-Zero Sampling has been rediscovered and widely used, *e.g.* in a field study by KUMMER (1965, 1968), in a study of caged cats by COLE & SHAFER

---

4) The terms "time samples" and "time sampling" have been variously used by different writers. SADE (1966) refers to his observability samples as *time samples.* In studies of human behavior, *time sampling* has been used by some to refer only to One-Zero Scoring, as described by GOODENOUGH, and by others to mean almost any sampling in which a fixed time unit of observation is maintained (see *e.g.* OLSEN & CUNNINGHAM, 1934; CONNOLLY & SMITH, 1972). In this paper I use the terms "time sample" and "time sampling" only when, for clarity, it seems advantageous to use the same term as a particular author in discussing that author's work — in which case these terms are used in quotation marks. I refer to SADE's samples simply as SADE's *observability samples,* to those techniques defined by GOODENOUGH as *One-Zero Sampling,* and I have labeled other sampling techniques utilizing a time base according to their distinctive features (*e.g.* Focal-Animal Sampling, Section V).

(1966), in studies of caged groups of monkeys by LINDBURG (1969), MENZEL (1963), BERNSTEIN & DRAPER (1964), BERNSTEIN (1968), and RHINE & KRONENWETTER (1972), by HINDE and his students (*e.g.* HINDE, 1964, 1967) and by HANSEN (1966) and others at the University of Wisconsin (*e.g.* MITCHELL, 1968a, 1968b; SEAY, 1966; SUOMI, *et al.*, 1971). This last group of investigators refer to the method as a "Hansen system". Recent examples in child behavior research include the work of HUTT (1966), and RICHARDS & BERNAL (1972).

It should be noted here that some workers (*e.g.* ARRINGTON, 1943; KUMMER, 1968) indicate that at some times (or for some behaviors) they actually recorded all occurrences; but their data was tabulated, compared, and presented as One-Zero Scores. The discussion that follows refers to One-Zero Scores, regardless of whether they result from the method of recording or of tabulating. I consider the technique in detail because of its widespread use in observational studies of humans and of caged animals, because of indications that it is now beginning to be used by a number of field primatologists, and because of my serious reservations about its value in most situations.

### B. INTERPRETATION OF SCORES

Authors that use One-Zero Scores usually state, for each individual or specified class of individuals, the number of sample intervals and the number (or percentage) of intervals that included at least one occurrence of the behavior(s) in question. These scores may be combined or averaged over several sample sessions or for several individuals. In some cases, the differences between scores are tested by means of non-parametric tests, such as the Mann Whitney U Test (see *e.g.* RICHARDS & BERNAL, 1972), or scores for different behaviors subjected to correlational analysis (see *e.g.* MITCHELL, 1968b).

Frequency of communicative acts and time spent in various states are two common variables measured in behavioral research, and are assumed to be important to the animals. Most users of One-Zero Scoring, implicitly or explicitly, seem to have assumed that these scores provide good measures of one or both of these variables. Examples of such use of these scores can be found in BISHOP (1951), HINDE & SPENCER-BOOTH (1964, 1967), in which the scores are treated as behavior frequencies, and in HINDE & SPENCER-BOOTH (1964, 1967), and KUMMER (1968), in which they are treated as representing percent of time spent in an activity. It is too easy for both author and reader to forget that a One-Zero Score is not the frequency of *behavior* but is the frequency of *intervals* that included any amount

of time spent in that behavior. Such lapses occur in SMITH and CONNOLLY's recent review of "time sampling" as well as in the papers just cited. Nor is the percentage of intervals the same as the percentage of time spent in an activity. In what follows, we shall examine the relationship between these scores and the frequency, duration, and proportion of time spent in specified activities.

Of the authors cited in this section, all who are explicit about their scoring method indicate that they scored states. From what I can infer, most others did so, too. However, it is conceivable that some observers would score events; in seminars, when I have discussed One-Zero Scoring of states, several people have suggested One-Zero event recording as a way of re-moving the defects of this scoring system. For that reason, I also discuss here the case in which the recording is of events.

State scores.

As I noted previously, most of the studies that have utilized One-Zero Scores have scored presence or absence of states, rather than of events. That is, an act that began in one interval, continued through a second, and terminated in a third would result in three scores, one for each of these intervals. But three occurrences, each with onset and termination in the same interval, would yield only one score for all three of them. Thus, there is in general no direct relationship between such scores and the true fre-quencies. However, SMITH & CONNOLLY (1972) suggest that under special conditions, One-Zero Sampling will provide data on frequency and durations of behaviors. They write:

> "If the time sample period is much less than the behaviour duration (bout length) then the distinction between frequency and all-or-none recording vanishes. Use of sequential samples gives information on both number of occurrences and durations."

What is assumed in such use is a one-to-one correspondence between any onset of behavior and the corresponding record for two consecutive intervals, in which the first, taken just before the onset, contains no score (zero) and the second contains a score (one). Likewise, it is assumed that a score (one) in one interval followed by the absence of one (zero) in the next interval bears a one-to-one correspondence to the termination of behavior. Another way of looking at it is that one assumes that only one onset or termination (but not both) of a behavior being scored can occur in one interval. The probability of more than one onset in an interval must be negligible, as must the probability of both an onset and a termination (either of the same occur-

rence of a behavior, or the termination of one occurrence and the onset of the next). This not only requires that the sample intervals be much shorter than the "usual" behavior duration, as SMITH and CONNOLLY suggest, but much shorter than the "usual" intervals between behaviors as well. How much shorter for any level of probability to be considered "negligible" will depend on the distributions of behavior onsets, of durations, and of the length of intervals between behaviors. Obtaining adequate information on these distributions would require extensive sampling by an unbiased method (such as those discussed in Sections V and VI). If adequate samples were then available, the One-Zero Samples would usually be superfluous.

Do these One-Zero Scores provide a good measure of percent of time spent in a given behavior? The precent of intervals containing a One-Zero Score is used by some researchers as a measure of the percent of time spent in a behavior. This would be correct only if the behavior in question took up all of the time in each interval in which it was scored, and none of the time in the others.

The percent of intervals including a score will be an upper bound on the percent of time spent. How close the true value is to the upper bound will depend on how much of a "scored" interval is in fact taken up on the average by the behavior. Clearly, the shorter the intervals, relative to the behavior durations, the closer this upper bound will be to the time spent. However, durations are likely to vary between individuals, over time for the same individuals, and from one behavior to another — the very classes which are usually being used for comparisons.

Thus differences in two such scores cannot be attributed to differences in the proportion of time spent in an activity (or to differences in frequencies of occurrence) unless it is known that the scores provide consistent measures of such. In the absence of that knowledge, providing an interpretation for the scores remains a problem.

A few workers recognize that these scores do not represent either frequencies or time spent. MITCHELL (1968b) addresses himself to this point:

> "It is emphasized here that a Hansen frequency is not a true frequency of occurrence. When it is stated that the Hansen system was utilized to measure visual orients of a mother toward her infant it is not meant that each and every glance at the infant was recorded. Only one visual orient was recorded whether the mother looked at her infant once during a fifteen second interval or several times during that interval. Since there are 60 fifteen second intervals in a fifteen minute test session, an upper limit has been imposed on the number of times a behavior can

occur. This procedure allows the experimenters to observe several
behaviors at a time without sacrificing observer reliability, but the
numbers which result reflect a little of both the duration and the fre-
quency of a behavioral act, not just frequency alone."

Thus, the resultant scores may be greater than, equal to, or less than, the
true frequencies. As MITCHELL has observed above, the numbers reflect a
little of both the duration and the frequency of behavior. However, that
seems to be a weakness of such scores, rather than an advantage: they do
not provide accurate information about either.

Event scores.

Would the interpretation become any more sound if events rather than
states were scored? First, could we determine or estimate the percent of
time spent in a behavioral state? Even total frequency records during the
sample intervals would not enable us to do so unless we also had information
about the distribution of durations of the behavioral events in question.
Therefore, One-Zero Scores could be used to estimate time spent only if
such information about durations was used in combination with a technique
for estimating the true frequencies from the One-Zero Scores.

Consider, then, how One-Zero Scores are related to frequency. If events
rather than states were sampled — that is, if an occurrence was recorded
only if the defining event for that behavior occurred in that interval —
then the "score", which is the number of sample periods with such a record
of occurrence, tells us that at least that many events occurred in that session.
That is, we would have a lower limit on the number of occurrences. But
surely, unless we know that this lower limit is close to (or bears a known
fixed relationship to) the true frequency of occurrence, such a lower limit
tells us little. The relationship between this lower limit and the true frequency
may vary from individual to individual or over time for one individual.

Is there any other way that we can utilize these scores to get at fre-
quencies? ALTMANN & WAGNER (1970) suggest that we look at the problem
as one of estimation of the rate of occurrence of the behaviors. The as-
sumption again is that *only events* have been scored. They write as follows:

"Suppose that the temporal distribution of the behavior can be described
by a Poisson process; we will return later to what this implies. If
so, then the probability $p_0$ of no occurrence of the behavior in an inter-
val of length T is given by

$$p_0 = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}, \qquad (1)$$

where $\lambda$ is the mean rate of occurrence of events, and e is the base of natural logarithms. From eq. (1) we have $\log_e p_0 = -\lambda t$, and thus $\lambda = (-\log_e p_0)/t$. The maximum likelihood estimate of $p_0$ is obtained from the number n of intervals in which the behavior did not occur divided by N, the total number of intervals. Thus, $\lambda$ can be estimated as follows:

$$\lambda = -\frac{\log_e (n/N)}{t} \qquad (2)"$$

ALTMANN and WAGNER point out that:

"Use of the Poisson distribution implies that the behavior occurs randomly at a constant rate, that the chance of two or more simultaneous occurrences of the behavior is negligible, and that the chance that a particular behavior will occur during an interval is independent of the time that has elapsed since the last occurrence of that behavior."

However, even if one feels confident that the data could be approximated by a Poisson distribution, the estimates obtained from such scores would usually not be as good as those obtained from the true frequency distribution: too many data have been discarded (see FIENBERG, 1972).

I have recently learned that this approach is used in estimation of density within bacterial samples and insect populations in which the spatial frequency distribution is assumed to be Poisson. Its use in that context apparently was first suggested by FISHER (1935).

If it is not reasonable to assume that the behavior has a Poisson distribution, and if the actual frequency distribution is not known and one cannot reasonably guess at those properties that could be used to relate the One-Zero Scores to the true frequencies, then there is no basis for a frequency interpretation of a One-Zero Score of events, and *a fortiori* no basis for using them to estimate percent of time spent in an activity.

### C. EASE OF SCORING AND OBSERVER AGREEMENT

One-Zero Scoring has been advocated on two other grounds: that such scoring is easier to do, and that greater observer agreement results. As for the first, the observer should ask himself whether the effort saved is worth the information lost. Even under field conditions, or in studies of human social groups, it may be possible to obtain complete frequency and duration records, instead of One-Zero Scores. CHAPPLE *et al.* (1963) did so in studies of patients on a psychiatric ward, and STRUHSAKER (1971) was able to do so for a study of infant vervet monkeys in their natural habitat in Africa,

while JENSEN and his colleagues (*e.g.* JENSEN *et al.* 1967) did so for caged mother-infant pairs of monkeys, even though several laboratory studies of monkey infant behavior and mother-infant interactions have resorted to One-Zero Scores. If ease of use is the deciding factor, other techniques often would be preferable, such as Instantaneous and Scan Sampling (Section IX) which is easier to do than One-Zero Sampling.

As for the second claim, it is true that two independent observers, watching the same behavior, may get One-Zero Scores that closely agree; but as I have indicated, such scores will not, in general, be an accurate indication of either the frequency or the proportion of time spent on a behavior. In fact, it seems the poorer that One-Zero Scores are as measures of these variables, the greater will be the observer agreement. Thus, if three maternal glances occurred in an interval, two observers will agree on a check for that interval as long as both saw at least one glance. But if exact frequencies were being recorded, they would agree only if both saw all three glances. By the same token, if a behavior took up, say, half an interval, the One-Zero recorders would agree as long as each thought it took up some part of the interval. Greater agreement does not guarantee more information.

In short, neither ease of use nor observer agreement *per se* provide an adequate justification for the use of this technique. Despite this, and despite the absence in such scores of reliable information about frequency and time spent, an observer might maintain that such scores are good predictors of other phenomena; this would have to be demonstrated in each case. For those who consider frequency and duration of behaviors and percent of time spent in various states as variables of interest, alternative sampling methods should be considered (see sections V and IX).

## IX. INSTANTANEOUS AND SCAN SAMPLING

Instantaneous Sampling is a technique in which the observer records an individual's current activity at preselected moments in time (*e.g.* every minute on the minute throughout the day). It is a sample of states, not events.

Such sampling has been used to study the percent of time spent in various activities by caged golden hamsters ( M. P. M. RICHARDS, 1966), squirrels (C. C. SMITH, 1968), adult male baboons (SLATKIN, unpublished), and humans (BINDRA & BLOND, 1958; SMITH & CONNOLLY, 1972, and presumably CONNOLLY & SMITH, 1972).

Instantaneous Sampling can be used to obtain data from a large number of group members, by observing each in turn. Moreover, if the behavior of all visible group (or subgroup) members are sampled within a very short time period the record approaches a simultaneous sample on all individuals.

We shall refer to such Instantaneous Sampling on groups as Scan Sampling. If such sampling is done frequently, data are obtained on the time distribution of behavioral states in the whole social group. In particular, data are obtained on behavioral synchrony in the group. Such data are almost impossible to obtain by most other sampling techniques (cf. Sections V and VI).

Such Scan Sampling has been used by COHEN (1971b) in his studies on subgroups of children, and by COHEN (1971a) and S. ALTMANN (unpublished) in studies of subgroups in yellow baboons and gelada monkeys, respectively, by CHALMERS (1968a, b) in his attempt to estimate differential visibility of monkeys (see Section III of this monograph), by CHALMERS (1968a) and by S. & J. ALTMANN (1970) to sample diurnal variation in activities and synchrony of activities in mangabey and baboon groups, respectively.

Censuses are essentially Scan Samples in which one obtains data on population parameters (such as age-sex distribution) as well as total group size. In censuses one is not usually concerned with approximating an instantaneous sample because the change from one age-state, for example, to another is quite long relative to the time necessary to complete the census.

## A. PRACTICAL CONSIDERATIONS

In an ideal Instantaneous Sample, each individual's state would be instantly noted. If, in addition, the state of the entire group is of interest (as in studies of subgroups or of synchrony), then ideally the state of every individual in the group would be noted at the same moment in time: the scan should be instantaneous. In practice, however, the observation, classification, and recording of a state takes time, and so does scanning from one individual to the next. The observer should try to scan each individual for the same brief period of time, for otherwise, a scan sample is equivalent to a series of short Focal-Animal samples of variable and unknown durations. In order to keep sampling time brief, the categories that are recorded should be easily and quickly distinguished. For this reason, it is in general more suited to studies of non-social behavior (with all social activities lumped into one state, as SLATKIN did in his study), or to situations in which social behaviors can be lumped into a few easily distinguished categories.

Of the reports cited above in which Scan Sampling was utilized, none indicate the time spent per individual, or whether an attempt was made to keep the times brief and even. Several indicate the amount of time per complete scan. COHEN (personal communication) took 5-10 seconds for his scans of subgroups in nursery school children, and about 45 seconds for a group

of savannah baboons. S. ALTMANN (personal communication) took three to seven minutes to scan and record subgroups involving up to 331 individuals in herds of gelada monkeys that ranged up to 425 individuals. S. & J. ALTMANN (unpublished) took 45-60 seconds for scans of a group of baboons (about 40 individuals) in which we noted behavior, rather than subgroup sizes; 60 seconds were required when we obtained both kinds of data (behavior and subgroup affiliations) for a group of about 38 baboons with about 30 individuals visible for each sample.

## B. ESTIMATING PERCENT OF TIME

A primary use of Instantaneous Sampling is in studies of the amount or percent of time that individuals devote to various activities. The percent of time is estimated from the percent of samples in which a given activity (state) was recorded. As I noted in the summary of One-Zero Sampling (p. 258), Instantaneous Sampling is at least as easy as One-Zero Sampling and, unlike One-Zero Sampling, readily provides data appropriate to estimating percent of time spent in various activities. In most of the studies cited above, Instantaneous Sampling was used to obtain such estimates. SMITH & CONNOLLY (1972, pp. 70-71) explicitly chose Instantaneous Sampling over One-Zero Sampling for this purpose.

SLATKIN was interested in baboon time budgets and considered it crucial for his purposes that he stay with an individual throughout a day. While Focal-Animal Sampling would have been the ideal method, he found it impossible to record behavioral and activity transition time data throughout an entire day. With the aid of an assistant, he did such Focal-Animal Sampling for selected half-hour periods (see p. 228); during the rest of the day he did Instantaneous Sampling at one-minute intervals.

## C. ESTIMATING RATES AND RELATIVE FREQUENCIES

Instantaneous Samples are discrete samples of states, *i.e.,* of ongoing behaviors. They are not samples of events, or transition times between states. It is true that under some sets of *ad hoc* assumptions about the distributions of the transition times, or of the durations of the states, it is possible to use Instantaneous Sample data to estimate transition rates, but without such assumptions such data by themselves provide no information whatever about rates of events or transitions — except of course that the number of consecutive samples exhibiting differing states does give a crude lower bound for the number of transitions. In the special case where the interval between Instantaneous Samples is short enough that no more than one transition can occur between consecutive samples, the resulting data are essentially

equivalent to that of Focal-Animal Sampling for rate and relative frequency estimates, but have a greater margin of error for duration estimates. (See the corresponding discussion for One-Zero Sampling, p. 254). However, utilizing such sufficiently short intervals will usually be no easier than Focal-Animal Sampling, while providing less information. If events, rates or relative frequencies are of primary importance in a study, the sampling method of choice would be one in which transition times, or other events, are recorded (see, *e.g.*, Focal-Animal Sampling, Section V). The necessary data are then provided directly.

## SUMMARY

Seven major types of sampling for observational studies of social behavior have been found in the literature. These methods differ considerably in their suitability for providing unbiased data of various kinds. Below is a summary of the major recommended uses of each technique:

| Sampling Method | State or Event Sampling | Recommended Uses |
|---|---|---|
| 1. *Ad Libitum* (p. 235) | either | Primarily of heuristic value; suggestive; records of rare but significant events. |
| 2. Sociometric Matrix Completion (p. 240) | event | Asymmetry within dyads. |
| 3. Focal-Animal (p. 242) | either | Sequential constraints; percent of time; rates; durations; nearest neighbor relationships. |
| 4. All Occurrences of Some Behaviors (p. 247) | usually event | Synchrony; rates. |
| 5. Sequence (p. 248) | either | Sequential constraints. |
| 6. One-Zero (p. 251) | usually state | None. |
| 7. Instantaneous and Scan (p. 258) | state | Percent of time; synchrony; subgroups. |

In this paper, I have tried to point out the major strengths and weaknesses of each sampling method.

Some methods are intrinsically biased with respect to many variables, others to fewer. In choosing a sampling method the main question is whether the procedure results in a biased sample of the variables under study. A method can produce a biased sample directly, as a result of intrinsic bias with respect to a study variable, or secondarily due to some degree of dependence (correlation) between the study variable and a directly-biased variable.

In order to choose a sampling technique, the observer needs to consider carefully the characteristics of behavior and social interactions that are relevant to the study population and the research questions at hand. In most studies one will not have adequate empirical knowledge of the dependencies between relevant variables. Under the circumstances,

the observer should avoid intrinsic biases to whatever extent possible, in particular those
that direcly affect the variables under study.

Finally, it will often be possible to use more than one sampling method in a study.
Such samples can be taken successively or, under favorable conditions, even concurrently.
For example, we have found it possible to take Instantaneous Samples of the identities
and distances of nearest neighbors of a focal individual at five or ten minute intervals
during Focal-Animal (behavior) Samples on that individual. Often during Focal-Animal
Sampling one can also record All Occurrences of Some Behaviors, for the whole social
group, for categories of conspicuous behavior, such as predation, intergroup contact,
drinking, and so on. The extent to which concurrent multiple sampling is feasible will
depend very much on the behavior categories and rate of occurrence, the observational
conditions, *etc.* Where feasible, such multiple sampling can greatly aid in the efficient
use of research time.

## BIBLIOGRAPHY

ALDRICH-BLAKE, F. P. G. (1970). Problems of social structure in forest monkeys. —
In: J. H. CROOK (ed.), Social Behaviour in Birds and Mammals. New York, Academic
Press.
ALEXANDER, B. K. & BOWERS, J. M. (1969). Social organization of a troop of Japanese
monkeys in a two-acre enclosure. — Folia primat. 10, p. 230-242.
ALTMANN, S. A. (1965). Sociobiology of rhesus monkeys. II: Stochastics of social
communication. — J. theoret. Biol. 8, p. 490-522.
—— & ALTMANN, J. (1970). Baboon Ecology. African Field Research. — Bibl. Primatol.,
No. 12. Basel, S. Karger and Chicago, Univ. Chicago Press.
—— & —— (In prep.) Baboon social behavior.
—— & WAGNER, S. S. (1970). Estimating rates of behavior from Hansen frequencies. —
Primates 11, p. 181-183.
ARRINGTON, R. E. (1943). Time sampling in studies of social behavior: a critical review of
techniques and results with research suggestions. — Psychol. Bull. 40, p. 81-124.
BARKER, R. G. (1963). The stream of behavior as an empirical problem. — In: R. G.
BARKER (ed.), The Stream of Behavior. New York, Appleton-Century-Crofts, p. 1-22.
BERNSTEIN, I. S. (1968). The lutong of Kuala Selangor. — Behaviour 32, p. 1-16.
—— & DRAPER, W. A. (1964). The behaviour of juvenile rhesus monkeys in groups. —
Anim. Behav. 12, p. 84-91.
BEER, C. G. (1961). Incubation and nest building behaviour of black-headed gulls. I:
Incubation behaviour in the incubation period. — Behaviour 18, p. 62-106.
—— (1962). Incubation and nest-building behaviour of black-headed gulls. II: Incubation
behaviour in the laying period. — Behaviour 19, p. 283-304.
—— (1963). Incubation and nest-building behaviour of black-headed gulls. III: The
pre-laying period. — Behaviour 21, p. 13-77.
—— (1963). Incubation and nest-building behaviour of black-headed gulls. IV: Nest-
building in the laying and incubation periods. — Behaviour 21, p. 155-176.
BINDRA, D. & BLOND, J. (1958). A time-sample method for measuring general activity
and its components. — Can. J. Psychol. 12, p. 74-76.
BISHOP, B. M. (1951). Mother-child interaction and the social behavior of children. —
Psychol. Monogr.: General and Applied, H. S. CONRAD (ed.), 65, p. 1-34.
BLURTON-JONES, N. G. (1967). An ethological study of some aspects of social behaviour
of children in nursery school. — In: D. MORRIS (ed.), Primate Ethology. Chicago,
Aldine, p. 347-368.
—— (1972). Categories of child-child interaction. — In: BLURTON-JONES, N. (ed.),
Ethological Studies of Child Behaviour. Cambridge, Cambridge University Press,
p. 97-127.

CHALMERS, N. R. (1968a). Group composition, ecology and daily activities of free living mangabeys in Uganda. — Folia primat. 8, p. 247-262.

—— (1968b). The social behaviour of free living mangabeys in Uganda. — Folia primat. 8, p. 263-281.

CHAPPLE, E. G., CHAMBERLAIN, A., ESSER, A. & HANDKLINE, N. S. (1963). The measurement of activity patterns of schizophrenic patients. — J. nerv. ment. Dis. 137, p. 258-267.

COCHRAN, W. G. (1954). Some methods for strengthening the common $\chi^2$ tests. Biometrics 10, p. 417-451.

COHEN, J. E. (1971a). Social grouping and troop size in yellow baboons. — Proc. 3rd Int. Congr. Primat. Vol. 3. Basel, S. Karger, p. 58-64.

—— (1971b). Casual Groups of Monkeys and Men; Stochastic Models of Elemental Social Systems. Cambridge, Harvard Univ. Press.

COLE, D. D. & SHAFER, J. N. (1966). A study of social dominance in cats. — Behaviour 28, p. 39-53.

CONNOLLY, K. & SMITH, P. K. (1972). Reactions of pre-school children to a strange observer. — In: BLURTON-JONES, N. (ed.), Ethological Studies of Child Behaviour. London, Cambridge Univ. Press, Ch. 6.

CRAIG, J. V., BISWAS, D. K. & GUHL, A. M. (1969). Agonistic behavior influenced by strangeness, crowding and heredity in female domestic fowl (Gallus gallus). — Anim. Behav. 17, p. 498-506.

DOYLE, G. A., ANDERSSON, A. & BEARDER, S. K. (1969). Maternal behaviour in the lesser bushbaby (Galago senegalensis moholi) under semi-natural conditions. — Folia primat. 11, p. 215-238.

FIENBERG, S. E. (1972). On the use of Hansen frequencies for estimating rates of behavior. — Primates, 13, p. 323-326.

FISHER, R. A. (1935). The Design of Experiments. — London, Oliver & Boyd.

FISLER. G. F. (1967). Nonbreeding activities of three adult males in a band of free-ranging rhesus monkeys. — J. Mammal. 48, p. 70-78.

GOODENOUGH, F. L. (1928). Measuring behavior traits by means of repeated short samples. — J. juv. Res. 12. p. 230-235.

HALL, K. R. L. (1965). Experiment and quantification in the study of baboon behavior in its natural habitat. — In: VAGTBORG, H. (ed.), The Baboon in Medical Research, p. 29-42.

HANSEN, E. W. (1966). The development of maternal and infant behavior in the rhesus monkey. — Behaviour, 27, p. 107-149.

HAUSFATER, G. (1971). Seasonal changes in participation in the intergroup encounters of free-ranging rhesus monkeys (Macaca mulatta). — Paper delivered at 40th annual meeting of the American Association of Physical Anthropologists, Boston.

—— (1972). Intergroup behaviour of free-ranging rhesus monkeys (Macaca mulatta). — Folia primat. 18, p. 78-107.

HAZLETT, B. A. & BOSSERT, W. H. (1965). A statistical analysis of the aggressive communications systems of some hermit crabs. — Anim. Behav., 13, p. 357-373.

HINDE, R. A., ROWELL, T. E. & SPENCER-BOOTH, Y. (1964). Behaviour of socially living rhesus monkeys in their first six months. — J. Zool. 143, p. 609-649.

—— & SPENCER-BOOTH, Y. (1968). The study of mother-infant interaction in captive group-living rhesus monkeys. — Proc. Roy. Soc., London, Ser. B, 169, p. 177-201.

HOPF, S. (1972). Study of spontaneous behavior in squirrel monkey groups: Observation techniques, recording devices, numerical evaluation and reliability tests. — Folia primat. 17, p. 363-388.

HUTT, C. (1966). Exploration and play in children. — Symp. Zool. Soc. London 18, p. 61-81.

JENSEN, G. D., BOBBITT, R. A. & GORDON, B. N. (1967). The development of mutual independence in mother-infant pigtailed monkeys, *Macaca nemestrina*. — In: S. A. ALTMANN (ed.), Social Communication Among Primates, p. 43-53.

KUMMER, H. (1968). Social Organization of Hamadryas Baboons. A field study. Chicago, University of Chicago Press and Basel, S. Karger, Bibl. Primatol. No. 6.

—— & KURT, F. (1965). A comparison of social behavior in captive and wild hamadryas baboons. — In: VAGTBORG (ed.), The Baboon in Medical Research. Austin, Univ. Texas Press, p. 65-80.

—— (1971). Primate Societies. Group Techniques of Ecological Adaptation. Chicago, Aldine-Atherton.

LINDBURG, D. G. (1969). Behavior of infant rhesus monkeys with thalidomide-induced malformations: A pilot study. — Psychon. Sci. 15, p. 55-56.

—— (1971). The rhesus monkey in North India: an ecological and behavioral study. — In: ROSENBLUM, L. A. (ed.), Primate Behavior. Developments in Field and Laboratory Research, 2, p. 1-106.

MASON, W. A. & EPPLE, G. (1969). Social organization in experimental groups of *Saimiri* and *Callicebus*. — In: C. R. CARPENTER, (ed.), Proc. 2nd Int. Congr. Primat., 1 (Behavior), Basel, S. Karger, p. 59-65.

McGREW, W. C. (1972). Aspects of social development in nursery school children with emphasis on introduction to the group. — In: N. BLURTON-JONES (ed.), Ethological Studies of Child Behaviour. London, Cambridge Univ. Press, Ch. 5.

MENZEL, E. W., Jr. (1963). The effects of cumulative experience on responses to novel objects in young isolation-reared chimpanzees. — Behaviour 21, p. 1-12.

—— (1969). Naturalistic and experimental approaches to primate behavior. — In: E. P. WILLEMS & H. RAUSH (eds.), Naturalistic Viewpoints in Psychological Research. New York, Holt, Rinehart & Winston, p. 78-121.

MISSAKIAN, E. A. (1972). Geneological and cross-genealogical dominance relations in a large group of free-ranging rhesus monkeys (*Macaca mulatta*) on Cayo Santiago. — Primates 13, p. 169-180.

MITCHELL, G. D. (1968a). Attachment differences in male and female infant monkeys. — Child Develop. 39, p. 611-620.

—— (1968b). Intercorrelations of maternal and infant behaviors in *Macaca mulatta*. — Primates 9, p. 85-92.

OLSON, W. C. (1929). The Measurement of Nervous Habits in Normal Children. — Minneapolis, Univ. Minnesota Press.

—— & CUNNINGHAM, E. M. (1934). Time-sampling techniques. — Child Dev. 5, p. 41-58.

PLOOG, D. W. (1967). The behavior of squirrel monkeys (*Saimiri sciureus*) as revealed by sociometry, bioacoustics, and brain stimulation. — In: S. A. ALTMANN (ed.), Social Communication Among Primates. Chicago, Univ. Chicago Press, p. 149-184.

——, BLITZ, J. & PLOOG, F. (1963). Studies on social and sexual behavior of the squirrel monkey (*Saimiri sciureus*). — Folia primat. 1, p. 29-66.

RHINE, R. J. & KRONENWETTER, C. (1972). Interaction patterns of two newly formed groups of stumptail macaques (*Macaca arctoides*). — Primates, 13, p. 19-33.

RICHARD, A. (1970). A comparative study of the activity patterns and behavior of *Alouatta villosa* and *Ateles geoffroyi*. — Folia primat. 12, p. 241-263.

RICHARDS, M. P. M. (1966). Activity measured by running wheels and observation during the oestrous cycle, pregnancy and pseudo-pregnancy in the golden hamster. — Anim. Behav. 14, p. 450-458.

—— & BERNAL, J. F. (1972). An observational study of mother-infant interaction. — In: N. BLURTON-JONES (ed.), Ethological Studies of Child Behaviour. London, Cambridge Univ. Press. Ch. 7.

ROSENBLUM, L. A. & KAUFMAN, I. C. (1967). Laboratory observations of early

mother-infant relations in pigtail and bonnet macaques. — In: S. A. ALTMANN (ed.), Social Communication Among Primates. Chicago, Univ. Chicago Press.

ROWELL, T. E. (1967). A quantitative comparison of the behaviour of a wild and a caged baboon group. — Anim. Behav. 15, p. 499-509.

—— (1968). Grooming by adult baboons in relation to reproductive cycles. — Anim. Behav. 16, p. 585-588.

SAAYMAN, G. S. (1972). Effects of ovarian hormones upon the sexual skin and mounting behaviour in the free-ranging chacma baboon (Papio ursinus). — Folia primat. 17, p. 297-303.

SADE, D. S. (1966). Ontogeny of social relations in a group of free-ranging rhesus monkeys (Macaca mulatta Zimmerman). — Dissertation, Univ. Calif., Berkeley.

SCHNEIRLA, T. C. (1950). The relationship between observation and experimentation in the field study of behavior. — Ann. N. Y. Acad. Sci. 51, p. 1022-1044.

SCRUTON, D. M. & HERBERT, J. (1970). The menstrual cycle and its effect on behaviour in the talapoin monkey (Miopithecus talapoin). — J. Zool. 162, p. 419-436.

SEAY, B. (1966). Maternal behavior in primiparous and multiparous rhesus monkeys. — Folia primat. 4, p. 146-168.

SINGH, S. D. & MANOCHA, S. N. (1966). Reactions of the rhesus monkey and the langur in novel situations. — Primates 7, p. 259-262.

SLATKIN, M. (in preparation). Activity patterns in two baboon species (Theropithecus gelada and Papio cynocephalus).

SMITH, C. C. (1968). The adaptive nature of social organization in the genus of tree squirrels Tamiasciurus. — Ecol. Monogrs 38, p. 31-63.

SMITH, Mapheus. (1931). A study of the unsupervised behavior of a group of institutional children. — Nashville, Marshall & Bruce Co.

—— (1933). A method of analyzing the interaction of children. — J. juv. Res., 17, p. 78-88.

SMITH, P. K. & CONNOLLY, K. (1972). Patterns of play and social interaction in pre-school children. — In: N. BLURTON-JONES (ed.), Ethological Studies of Child Behaviour. London, Cambridge Univ. Press. Ch. 3.

STRUHSAKER, T. T. (1967). Behavior of elk (Cervus canadensis) during the rut. — Zs. f. Tierpsychol. 24, p. 80-114.

—— (1971). Social behaviour of mother and infant vervet monkeys (Cercopithecus aethiops). — Anim. Behav. 19, p. 233-250.

SUOMI, S. J., HARLOW, H. F. & KIMBALL, S. D. (1971). Behavioral effects of prolonged partial social isolation in the rhesus monkey. — Psychol. Rep. 29, p. 1171-1177.

SWAN, C. (1938). Individual differences in the facial expressive behavior of pre-school children: a study by the time-sampling method. — Genet. Psychol. Monogr. 20, p. 557-650.

WASHBURN, R. W. (1932). A scheme for grading the reactions of children in a new social situation. — J. genet. Psychol. 40, p. 84-99.

WEBB, E. J., CAMPBELL, D. T., SCHWARTZ, R. D. and SECHREST, L. (1966). Unobtrusive Measures: Nonreactive Research in the Social Sciences. — Chicago, Rand McNally.

WOOTTON, R. J. (1972). The behaviour of the male three-spined stickleback in a natural situation: A quantitative description. — Behaviour 41, p. 232-241.

WRIGHT, H. F. (1960). Observational child study. — In: P. H. MUSSEN (ed.), Handbook of Research Methods in Child Development. New York, J. Wiley & Sons, p. 71-139.

ZUSAMMENFASSUNG

In der Literatur finden sich hauptsächlich sieben Methoden, um Beobachtungen vom Sozialverhalten zu erheben. Sie leisten objektiv Unterschiedliches und sind daher nicht für jeden Untersuchungsweck gleich geeignet. Die Tabelle faßt diese Unterschiede zusammen:

| Erhebungsmethode | Aufzeichnung: Verlaufsdauer oder -wechsel | Anwendungsbereich |
| --- | --- | --- |
| 1. Je nach Gelegenheit (S. 235) | beides | Hauptwert liegt in erster Bestandsaufnahme und im Auffinden von Hinweisen für die Anlage einer mehr systematischen Untersuchung; Entdeckung sehr seltener, aber theoretisch wichtiger Verhaltensweisen. |
| 2. Soziometrische Matrixausfüllung (S. 240) | -wechsel | Asymmetrie der Wechselbeziehungen von Zweiergruppen |
| 3. Konzentration der Beobachtung auf jeweils ein bestimmtes Tier (Fokus-Tier) (S. 242) | beides | Verlaufsregeln, % Zeit je Verhaltensweise; Dauer des Einzelablaufs; Häufigkeit; Verhalten zum nächsten Gruppenmitglied |
| 4. Aufzeichnung jedes Auftretens einer bestimmten Verhaltensweise in der Gruppe (S. 247) | meist: -wechsel | Synchronisierung; Häufigkeit |
| 5. Aufeinanderfolge von Verhaltensweisen (S. 248) | beides | Feststellung von Verlaufsordnungen und -regeln |
| 6. Auftreten oder Nichtauftreten von Verhaltensweisen in festgelegten Beobachtungsintervallen (S. 251) | meist: -dauer | nutzlos |
| 7. Gesamtfeststellung aller gleichzeitig ablaufenden Aktivitäten (S. 258) | -dauer | Synchronisierung; Zeitaufwand je Verhaltensweise; Untergruppen |

In dieser Zusammenstellung habe ich versucht, die hauptsächlichen Stärken und Schwächen jeder Erhebungsmethode darzulegen. Ob sich in einer gegebenen Forschungssituation eine Fehlerquelle erheblich oder nur unwesentlich auswirkt, hängt von dieser Situation und nicht von der gewählten Methode ab.

Allerdings sind manche Methoden hinsichtlich einer größeren Anzahl von Variablen mehr fehlerbelastet als andere. Wenn man eine bestimmte Erhebungsmethode auswählt, ist daher die erste Frage, ob die Methode gegenüber jenen Variablen zu fehlerhaften Ergebnissen führen könne, die man zu untersuchen wünscht, oder ob ein solcher Fehlereinfluß indirekt eintreten kann, weil eine der untersuchten Variablen in Korrelation mit einer anderen steht, die ihrerseits von der Methode nur unsicher erfaßt wird.

Bei der Auswahl der Erhebungsmethode muß sich der Untersucher sorgfältig überlegen, welche besonderen Verhaltenseigentümlichkeiten und gegenseitigen Beziehungen der Gruppenmitglieder sowohl für die untersuchte Population wie die besondere Fragestellung von Bedeutung sind. In den meisten Fällen wird man empirisch nicht genügend über die wechselseitige Abhängigkeit der zu untersuchenden Variablen wissen. Gerade dann soll sich der Untersucher bemühen, immanente Fehlerquellen möglichst auszuschließen, ganz besonders solche, welche die zu untersuchenden Variablen unmittelbar betreffen.

Schließlich läßt sich in einer Untersuchung oft mehr als nur eine Methode anwenden, nacheinander wie auch — unter günstigen Bedingungen — gleichzeitig. Z.B. konnten wir gleichzeitig (Methode 7) in Abständen von 5-10 Minuten Identität und Individualdistanzen der sich in der Nähe aufhaltenden Gruppenmitglieder feststellen, während wir ständig ein bestimmtes Tier im Auge hielten (Methode 3). Ebenso kann man oft, während man sich auf ein Tier konzentriert (Methode 3), zugleich jedes Auftreten bestimmter, besonders auffallender Verhaltensweisen (Beutefang, Trinken, Verhalten zu fremden Gruppen usw.) innerhalb der ganzen Gruppe mitaufnehmen (Mehode 4). Das Ausmaß, bis zu dem man mehrere Methoden nebeneinander verwenden kann, hängt natürlich von den zu beobachtenden Verhaltensweisen, ihrer Häufigkeit, den allgemeinen Beobachtungsbedingungen und ähnlichem ab. Die gleichzeitige Anwendung verschiedener Methoden kann — soweit dies möglich ist — dazu beitragen, die für die Untersuchung verfügbare Zeit wesentlich besser als sonst auszunutzen.

---