3 Observing Interactions and Associations: Collecting Data

3.1 Types of Behavior

Having decided what we wish to study and why (Chapter 1), the next step in analyzing social structure, and the subject of this chapter, is the collection of data. The real-life behavior of individual animals needs to be abstracted into a form, usually a numerical form, in which it can be analyzed. So, how should behavior be described? This subject is considered in detail by several authors, for instance, Lehner (1998, pp. 109–124) and Martin and Bateson (2007, pp. 48–61), and so in this chapter I summarize areas well covered in these books, concentrating on the most significant issues for social analysis.

There are two general ways of describing behavior—in terms of either the structure of the behavior (physical form, posture, movements, etc.) or the consequences of the structure for the animal or the environment (Martin & Bateson 2007, p. 32; Lehner 1998, p. 81).[1] Consequences might include agonism, feeding, or grooming. Describing behavior in terms of consequences is generally more powerful and economical (Martin & Bateson 2007, p. 32), but sometimes when we do not understand the behavior of

1. Lehner calls these types "empirical" and "functional" descriptions, respectively.

the animals well, it might be better to stay with structural definitions of behavior types. An example is when cetaceans are viewed from the surface and a characteristic posture of infants seems to suggest suckling, but nipple contact cannot be observed and we are not certain about the function of the behavior (Gero & Whitehead 2007).

Another important distinction is between behavioral events and behavioral states (Altmann 1974). Events are virtually instantaneous, and one must watch for awhile to see any. In any time period, they can be counted and classified. Events include behavior such as vocalizations, sudden movements, and ingestion of prey. States are behavior patterns of relatively long duration. Often, several states are defined [e.g., foraging, traveling, resting and socializing (Mann 2000)] such that an animal is always in at least one of them. They may be defined as exclusive, so that an animal is only in one state at a time, or be allowed to overlap (e.g., social and traveling). Events occur within states, and can be used to define them: The state "feeding" might consist of time periods that include events of food ingestion. Before starting a study, events and states should be defined as rigorously as possible, perhaps with the help of a preliminary study whose data are not used in the final analysis (Martin & Bateson 2007, pp. 31–32). If behavior is videotaped or recorded acoustically, then some of these decisions can be postponed, but rigor is important.

In this book, I am concerned with social structure, and therefore social behavior, which involves two or more individuals. The fundamental element of social behavior is the interaction (Section 1.6), an event that can be defined structurally or, more usually, by consequence. However, we can also use interactive state measures, associations, as the core of our social analysis or combine interactions and associations. Interactions and associations are the subjects of the next two sections, which are followed by a section on groups. Groups can be used as shortcuts to the designation of associations. All of this depends crucially on the ability to identify individuals (Section 3.5). Thus, the meat of most analyses of social structure comprises systematically collected records of observations of interactions, associations, or group membership among identified individuals. These may be supplemented by records of rare but important behavior and supplementary data about the individuals such as sex, age, reproductive status, or kinship (Section 3.6). The choice of sampling scheme depends on whether we are collecting interaction or association data, and other factors. These are considered in Section 3.7. The final sections of this chapter consider the formatting, structuring, and power of data sets.

3.2 *Interactions*

Interactions are the basis of Hinde's (1976) framework for the study of social structure. The definition of interaction is fairly straightforward ("the behavior of one animal is affected by the presence or behavior of another"), but the operational use of the term is less so (Hinde 1976). Is a fight one interaction, or are all the elements (physical contacts, vocalizations, movements, wounds, etc.) individual interactions? Do we separate different types of affiliative behavior or simply have one interaction type that may include affiliative grooming and vocalizations as well as movements?

My advice is to collect data at the highest level of resolution that permits operational consistency, so that interaction types are rarely missed or misclassified. If high operational resolution is feasible, it may turn out that some distinctly defined interaction types are functionally equivalent to the animals (interaction types A and B are interchangeable from their perspective) or redundant (A always follows B). A well-organized analysis will identify such relationships between interaction types and account for their interdependence in subsequent analyses (e.g., using principal components analysis; Section 2.6). Nothing is lost by collecting the high-resolution data, even if they are somewhat redundant. More frequently, especially with hard-to-study species, achievable operational precision will be much less than that which is meaningful to the animals. In these cases, the more detailed the data, the more meaningful is the model of social structure that results, provided that the data are reliable. For some analyses, however, it may be efficient to lump types of interaction into classes based on structure or function, such as "vocalizations" or "agonistic behavior." Although high-resolution data can be lumped if desired, it is not possible to go the other way, increasing resolution.

Sometimes, interactions are identified by synchrony or leader/follower events—one animal follows another in performing a particular activity—in other words, by temporal patterning (e.g., Connor et al. 2006). To identify synchrony or leader/follower events rigorously, we need simultaneous records of behavioral events from several individuals.

To give an idea of the range of types of interactions that can be recorded, Table 3.1 lists some, noting whether they are defined by structure or consequence (Section 3.1) and whether they are *symmetric* (if A interacts with B, then B interacts with A).

In what follows, I usually assume that interactions are dyadic—involving two animals—although some of the analyses make sense with triadic or higher-order interactions (e.g., Kummer et al. 1974).

Table 3.1 Examples of Interaction Types That Can Be Used as the Basis of a Social Analysis Together with Whether They Are Primarily Structural or a Consequence, or Primarily Symmetric or Asymmetric

Type	Structural/consequence	Symmetric?
Grooming bout	Consequence	No
Fight outcome	Consequence	No
Touch	Consequence	Yes
Synchronous dive	Consequence	Yes
Leader/follow dive	Consequence	No
Vocalization exchange	Consequence	Yes
Suckle	Structural	No
Intromission	Structural	No
Mating	Consequence	Yes
Particular gesture in dyadic context	Structural	No

3.3 Associations

Behavioral interactions, the foundation of Hinde's conceptual framework of social analysis, are events. Interactions cannot always be observed, however, and in some more cryptic species are virtually always hidden. A common way around this difficulty is to use *associations* instead of interactions or in addition to interactions as the fundamental elements of social analysis. Dyads are in "association" if they are in a situation in which interactions usually take place (Whitehead & Dufault 1999). Associations are state measures, and usually they are more easily measured than interactions. They can often be determined from nearly instantaneous observations, whereas interactions, even when observable, require prolonged observation.

Association can also be reasonably interpreted as "within range of communication" because communication involves the active or passive transmission of information that may change the behavior of the recipient (Bradbury & Vehrencamp 1998, p. 2), resulting in an interaction. This emphasizes the important role of communication in studies of social structure (Costa & Fitzgerald 1996).

Thus, ideally, the social analyst would initially make a thorough study of communication in her study species or have access to the results of one. From this, she could determine the dyadic circumstances that best characterize communication between a pair of animals. She can then define association in such a way that it delineates circumstances under which communication, and interactions, take place. Systematic records of such associations are then used as the data for social analysis. Such rigorous approaches are rare. Instead, even the best studies rarely go beyond reasoning such as the following: Animals can hear each other at

Table 3.2 Examples of Definitions of Dyadic Association That Can Be Used as the Basis of a Social Analysis Together with Whether They are Symmetric or Asymmetric

Definition	Symmetric?
Within x body lengths (or x meters)	Yes
Within x body lengths (or x meters), and in same behavioral state	Yes
Within x body lengths (or x meters), and heading in same direction	Yes
Nearest neighbor	No
Sharing feeding site/nest/roost	Yes
Duetting	Yes
Grooming	No
Overlapping home ranges	Yes
Grouped (table 3.3)	Yes

ranges up to about x meters, and so we will define association as "dyads separated by less than x meters." This is not unreasonable, and in most cases, errors in choosing a suitable x will not profoundly affect the subsequent analysis. A too small value of x will omit some interactions, and a too large one will include noninteracting dyads, but if a large data set is collected and there is no systematic bias (such as might be caused by pairs of individuals who generally interact at ranges just greater than the chosen x), an informative social model should emerge.

Association is usually defined based on spatial proximity plus, perhaps, some behavioral state measure (e.g., "within x body lengths and heading the same direction"). It is often possible, and desirable, to measure more than one association measure simultaneously. Perry (1996), for example, noted associations within 1, 5, and 10 body lengths for capuchin monkeys (*Cebus capucinus*); animals within 10 body lengths may interact vocally, and those within 1 body length may interact using touch. Table 3.2 lists some ways in which association has been defined.

Associations may be asymmetric. For instance, nearest neighbor is a commonly used asymmetric definition of association. During an observation, A may be the nearest neighbor of B, but B is not necessarily the nearest neighbor of A. Most analytical techniques assume symmetric association measures, so I generally recommend their use, although it should not take too much work to adapt them for asymmetric measures.

Sometimes, the same behavior can be viewed as either an interaction (event) or an association (state). The difference is whether the observations are considered continuous or instantaneous. For instance, we can count the number of grooming bouts in an observation period, considering each bout as an interaction, or observe whether a dyad is engaged in grooming behavior when we observe them, in which case grooming is an asymmetric association measure.

Association within any sampling period (Section 3.9) is usually defined in a 1:0 manner; a dyad is observed associated or is not observed associated within that sampling period. It is possible to define association as a continuous variable [e.g., a function of proximity or dive synchrony within the sampling period (Whitehead & Arnbom 1987; Perry 1996)], but this adds considerably to the complexity of the subsequent analysis, and I do not usually recommend this.

3.4 Groups

A simplifying assumption that is often made is that all individuals within some spatiotemporal "group" are associated with each other. Such groups are assumed to be *transitive* in the sense that if, at any time, A and B are members of the same group and so associated, and B and C are members of the same group and so associated, then A and C are also members of the same group and associated. We have called the assumption that grouped individuals are associated "the gambit of the group" (Whitehead & Dufault 1999). The fundamental assumptions, almost never tested, are that all, or almost all, interactions of some type take place within groups and that interactions of this type are similar and occur at a similar rate among all animals within a group (Whitehead & Dufault 1999).

To evaluate the likely validity of the gambit of the group, it helps to consider the reasons that animals may be in spatiotemporal proximity. An important distinction is between spatiotemporal clusters of individuals that are entirely the result of some nonsocial forcing factor, such as a localized source of food or shelter, and those that result from the active behavior of individuals converging on, or maintaining proximity with, other animals. I call these *aggregations* and groups, respectively, and only groups are of direct interest as elements of social structure. Connor (2000), making the assumption that animals do not usually behave maladaptively, refers to them as nonmutualistic groups (= aggregations) and mutualistic groups (= groups) because individuals are likely to seek or maintain proximity with other individuals if and only if there is expected to be mutual benefit. This suggests two ways of distinguishing aggregations from groups. They are groups if it can be shown either that individuals actively seek or maintain proximity with other individuals or that there is some benefit of being grouped with others.

In some cases, these criteria allow the simple recognition of groups. For instance, individuals clustered over a habitat with uniform resources (flocks of roosting birds in some trees when other nearby and similar trees are empty, or ungulates migrating over featureless habitat in groups) can be considered groups, as may clusters of individuals that are passed over by a predator that takes lone individuals. In other cases, it is not so clear whether clustered individuals form a group. Bats may be clustered in a cave solely because it is suitable roosting habitat or because they are drawn to assemblages of other roosting bats. Of course, even within my definition of group there is enormous variability. A fish school may contain animals drawn together solely to combat predation and possess little temporal stability and no behavioral substructure. In contrast, a nearby pod of killer whales (*Orcinus orca*) may be made up of genetic relatives who spend their lives together, feed together, and have distinctive relationships with each other (Ford et al. 2000). Such distinctions are considered in the following chapters, but at this step, we need to exclude aggregations from further social analysis.

Groups are usually obvious. If the spatiotemporal clustering is so subtle that human observers cannot be sure that it is present or so variable in type that they find it difficult to come up with a rigorous definition, then they should doubt whether such groups are meaningful to the animals. In many cases, however, although group distinctions are generally obvious, there are some borderline cases ("Is that one group or two?"). It is important that such instances be treated consistently, and so we need a criterion for allocating individuals to groups (Martin & Bateson 2007, pp. 46–47).

Ideally, we would base our definition of group on studies of communication, as with association (Section 3.3), but this is rarely done explicitly. In some fortunate instances, groups consist of sets of animals using small areas or volumes of suitable habitat separated clearly from other such groups. Roosting sites of birds and bats in particular trees, leaves, or caves are such cases (e.g., Vonhof et al. 2004), as are islets used by seals to haul out (e.g., da Silva & Terhune 1988). More usually, groups are formed over homogeneous or continuously varying habitat. A useful empirical approach is to measure interindividual distances, perhaps on photographs or video, and examine their distribution. If clear modes are apparent, then these can suggest a suitable definition of a group (Clutton-Brock et al. 1982, pp. 319–320). For instance, if there are many interindividual distances between 1 and 5 body lengths but very few between 6 and 30, then perhaps animals within 5 body lengths should be considered grouped. Whereas such a distance criterion works well with associations (Table 3.2), however, it may give inconsistent results when used directly to define groups: A and B may meet the criterion, and so may B and C, but A and C may not. In this case, the transitive feature of groups is violated. Frequently, researchers use a

Table 3.3 Examples of Criteria Used to Distinguish Groups

Definition
Sharing feeding site/nest/roost/haulout
Within *x* body lengths (or *x* meters) chain rule
Within *x* body lengths (or *x* meters) chain rule, and in same behavioral state
Within *x* body lengths (or *x* meters) chain rule, and coordinating movement
Clusters produced by *k*th nearest-neighbor clustering on spatial arrangement of individuals (Strauss 2001)

"chain rule" to circumvent this problem (Clutton-Brock et al. 1982, pp. 319–320; Smolker et al. 1992). If A and B meet the criterion as well as B and C, then A and C automatically meet it too, and so we have transitive groups.

Strauss (2001) considers the difficult case of fluid shoals of fish, which may form irregularly shaped groups. After some experimentation, he found that a method called "*k*th nearest-neighbor hierarchical clustering" best mimicked human perceptions of shoal membership. He also developed permutation methods to test whether the spatial arrangement of the animals is clustered compared with a random null hypothesis.

Table 3.3 list some criteria used to designate groups. Group criteria may include a behavioral condition such as "coordinated movement" or "in same behavioral state," which helps to exclude individuals that are incidentally clustered (Mann 2000). As with associations, it is possible simultaneously to use two or more types of group at different spatial or temporal scales, usually with one type nested within the other, which thereby stand in for different classes of interaction.

Estimating group size is usually easy if individuals can be assigned to groups: We just count. In some cases, however, for instance, with cryptic animals, a population assessment technique may be useful for estimating group size. Such techniques are summarized in Appendix 9.5.

3.4.1: Typical Group Size. A distinction that is important in many cases is between mean group size and mean *typical group size* (Jarman 1974). The former is the mean size as experienced by an outside observer, such as a predator or the social analyst, the latter is the size as experienced by a member of the population. For example, if there are four groups of size 1, 2, 2, and 3, then the mean group size is 2, whereas the mean typical group size is 2.25. There is 1 animal in a group of size 1, 4 in groups of size 2, and 3 in groups of size 3, giving a mean typical group size of $(1 \times 1 + 4 \times 2 + 3 \times 3)/8 = 2.25$. Mean typical group sizes are usually higher, and never lower, than those experienced by outside observers.

If we have counts or estimates of the sizes of n observed groups $N_g(i = i, \ldots, M)$ their mean is simply $\Sigma N_g(i)/M$, whereas the mean typical group size is $\Sigma N_g(i)^2/\Sigma N_g(i)$. We can also speak of the mean typical sizes of other social or nonsocial entities, such as aggregations, units, or communities. As an example of the differences between an animal- and an observer-centered approach, here are the mean sizes and mean typical sizes of aggregations (probably a nonsocial assemblage), groups, and social units (both social assemblages) of sperm whales (*Physeter macrocephalus*) (Whitehead 2003, pp. 213, 218):

	Mean	Typical mean
Aggregation (Ecuador 1991)	60.1	77.4
Group (South Pacific)	19.4	25.1
Unit (Galápagos)	10.5	13.6

3.5 Identifying Individuals

A primary requirement for the social analyses described in the subsequent chapters is that individuals be identifiable. There are many ways to do this. With humans and some small populations of other species, it may be possible to discriminate visually all individuals reliably in real time. In larger populations, photographs of natural markings can be used either to ground-truth visual identifications or as the sole source of identifications (Pennycuick 1978; International Whaling Commission 1990; Lehner 1998, pp. 221–223). Vision is not the only medium, however, by which individuals can be distinguished. In appropriate circumstances, individual identification may be possible using vocalizations (Adi et al. 2004) or by collecting DNA samples (Palsbøll et al. 1997; Sloane et al. 2000). Artificial marks, including dye marks and tags, can be used to identify individuals visually (Stonehouse 1978), acoustically (Zeller 1999), or through radio signals (Chambers et al. 2000). Use of passive integrated transponder (PIT) tags that can be implanted into animals or attached to them and "read" by an external radio signal is becoming an increasingly important method of identifying animals (Biomark, Inc., see http://www.biomark.com/; e.g., McCormick & Smith 2004).

In many situations, especially if tags or artificial marks are used, but also sometimes with naturally marked animals (e.g., Ottensmeyer & Whitehead 2003), only a portion of the population will be identifiable. This should be considered in subsequent social analyses. For instance, estimates of group sizes need to be corrected for animals that are not identified. In the worst scenario, the social behavior of the identified animals is not representative of the population. This could happen if more-aggressive individuals were more likely to accumulate natural markings or those generally in smaller groups were more likely to be tagged.

Another important issue with individual identification is reliability. With almost all techniques, there is a chance that an animal can lose or change the identifying feature. Natural marks can change, tags may be lost, or the technology in acoustic or radio tags may malfunction. In most cases, a numerical analysis treats this as equivalent to mortality or permanent emigration. Thus, if marking failure is likely, then methods and models that include the disappearance of the animal from the population, by whatever means, need to be used. If "mortality" or "survival" estimates are produced (as in some models of lagged association rates; Section 5.5), then it must be recognized that these include mark failure, as well as perhaps permanent emigration from the population. With natural markings, there is the additional possibility that an unrecognizable mark change will produce a "new" individual in the population, for instance, when a large, new mark obliterates the features previously associated with an individual or a "clean" unmarked individual gains marks and joins the study population. Although this is an important concern when estimating populations using natural markings (Hammond 1986), these changes, as long as they are not too frequent, are unlikely seriously to affect social analyses.

Misidentification is of greater concern. This can occur with all techniques and at several stages of the identification process. Real-time visual identifications can be wrong, acoustic identifications may be less than 100% accurate (Adi et al. 2004), and equipment problems can cause errors with acoustic or radio tags. The data can be recorded wrongly, entered in the database erroneously, or scrambled by a computer. All of these can cause data to enter the numerical analyses tagged with the wrong individual. With large data sets, there are almost certain to be some errors. One hopes that the rate will be small, but what will be the effect on the output measures of social structure? I cannot give an overall prescription, and each situation should be considered on its merits. The greatest danger, however, occurs when a particular data record can have a considerable effect on the output. This might be the case if we are looking for closed units, when a misidentification could mean rejecting this hypothesis or lumping separate units. With less hard-and-fast social structures, however, such as "fission–fusion societies," a small proportion of misidentifications is unlikely to have any major impact on the results. One can test the effects of misidentifications by deliberately making some and seeing how the output is changed. This method will not work, however, if errors have already ruled out a "true" model for the data, such as closed units.

3.6 Class Data

An important part of Hinde's (1976) framework (Fig. 1.4) is the generalization from interactions between two individuals to interactions between classes of individuals. This is often done by some form of averaging ("the mean rate at which mothers groom their neonates") or by abstracting essential characteristics of interrelationships between classes ("fights only occur between mature males") (Hinde 1976). These class-abstracted results are clearly of interest in their own right. When data are few or sparse, however, there may be insufficient power to categorize social structure at the level of individuals, and so class abstractions become the principal results of a social analysis (Section 4.11). Classes can also be used to form nonsocial measures of relationship (e.g., "same or different gender"; Section 4.2) that are important when trying to address functional questions (Chapter 7).

Thus, social analyses are much richer if animals can be classified using attributes of individuals (Table 1.2). Here are some of the classes most frequently used:

- *Sex*. Gender can be determined by observation, photographs or video of genital areas or sexually dimorphic anatomical features; observations of gender-specific behavior, such as nursing; or sex-specific DNA markers in tissue samples.
- *Age*. Ideally, this is available from the lifetime knowledge of an individual, but sometimes accurate aging of living animals can be achieved through other means, such as drawing and sectioning a tooth. Age can often be estimated by size. With inaccurate aging methods, it may be more appropriate to assign animals to general age classes.
- *Physiological state*. Classes may describe sexually mature or immature animals, pregnant or estrous animals, or some other physiological state.
- *Subspecies, morph, and so on*. These can be considered in mixed populations.
- *Matriline* (or *patriline*). These may be used in populations with well-known genealogies.

- *Genes.* The genetic class that is most usually employed is the mitochondrial haplotype (e.g., Weinrich et al. 2006), which in some respects stands in for the matriline when this is not known.
- *Behavioral phenotype.* Individuals can be classified into those that are dominant, aggressive, submissive, and so on (see Section 4.3).
- *Social unit, community.* Sometimes, the results of one level of social analysis can be used to define classes (Section 5.7) and these used to investigate questions such as whether patterns of within-unit or within-community social structure differ among units or communities or whether there are consistent affiliations among units.

3.7 Collecting Social Data

The analysis of social structure needs data. Social data may be recorded by human observers on data sheets, voice recordings, or photographs or keyed straight into computers (Section 2.9). Alternatively or additionally, acoustic, visual, or electronic data may be recorded by automated devices.

From the perspective of Hinde's framework and this book, these data are in the form of records of interactions and/or associations among identified individuals. Suppose a researcher is interested in a population of animals some or all of which are identifiable, and that she has a time frame over which the study is to be carried out and an effort budget. How should she plan data collection to give the most informative model of social structure? This is not an easy decision. It depends on the actual social structure, what behavior is observable, and a number of other factors.

Altmann's (1974) paper "Observational study of behavior: sampling methods" has been very influential in guiding the collection of behavioral data and has formed the basis of several other good reviews of protocols and procedures in behavioral observation (e.g., Martin & Bateson 2007, pp. 48–61; Lehner 1998, pp. 189–210; Mann 2000). Social data form a subset of behavioral data, and a subset with special characteristics, because two (or possibly more) individuals are involved. Thus, Altmann's (1974) recommendations need some refinement. In the following subsections, I consider whether to collect interaction, association, or group data and how to collect them. Table 3.4 summarizes the major recommendations and contrasts the features of interaction and association data.

Table 3.4 Collecting Interaction or Association Data: Guidelines

	Interactions	Associations
Type of measure	Event	State
Dyadic measure	Usually counts	Usually 1:0 (associated:not associated)
Use when	Interactions reliably and frequently observable	Interactions not reliably or frequently observable; coordinated behavior predominates
Follow protocol	Usually individual follow is best	Usually survey or group follow is best
Sampling protocol	All interactions involving focal animal with times and interactant identities is best	Associations are noted at regular times or when they change

3.7.1: Interactions, Associations, or Groups? As indicated in preceding sections, social analysis can be based on interactions, association, or groups. But which is preferable?

If interactions form the basis of social structure and associations are merely the imperfectly defined circumstances under which interactions are likely to occur (Section 3.3), then would it not be true that interactions should be preferred over associations as the targets of data collection? Under this rationale, it should be better to record when animals touch than that they are grouped, and associations should only be used as a "stand in" when interactions are unobservable. This used to be my perspective (Whitehead & Dufault 1999). After more reflection, however, I am not so sure (Whitehead 2004). A pair may have an important relationship but not touch or perform any overt interaction. A seamless behavioral synchrony without any observable interactions, as is sometimes characteristic of dolphins (Connor et al. 2006), might indicate the strongest of relationships. From the practical perspective, if we can rarely see animals interacting, associations will be more appropriate measures of sociality. In such circumstances, records of associations may be much more revealing than those of interactions.

In many circumstances, either associations or groups can be recorded. To decide which is preferable, we would ideally need to see inside the minds of the animals. Are their locations, movements, and behaviors more the result of the locations, movements, and behaviors of particular companions or of "the group" itself? Resolving this would often be a major study in its own right, but aspects of the animals' behavior can help. When movement is coordinated within a whole group or there are frequent changes of position within the group such as occurs in some fish schools, then the group appears to be a more significant behavioral determinant than the identity of any associate. In contrast, if behaviors and movements vary among clustered animals or they show little active behavior or movement, as in resting lions (*Panthera leo*), then perhaps associations should be used. If accurate positions are recorded, then both are possible, associations being defined by, for instance, one of the measures in Table 3.2 and then groups formed from the associations using a "chain rule" (Section 3.4). Group memberships are usually easier to record than associations, however, and so practical limitations may play a role in this decision.

There is no theoretical problem with recording interactions as well as associations and perhaps also groups (potentially derived from the association data), and more than one type of association, interaction, or group. If the different dyadic measures are well correlated and add little information, then methods such as principal components analysis can be used to simplify the multivariate data set (Section 5.6). If they are not well correlated, then the analysis is that much the richer.

3.7.2: Temporal Patterning and Length of Observations. Let us suppose that we have start and stop dates for the research, say, 1 June to 10 September, and also a limit on the total amount of research effort, say, 100 hours of observation. How should it be allocated? There are many possibilities, such as 1 hour per day every day for 100 days, or ten 10-hour days on 1 to 5 June and 26 to 31 August.

To make the best use of time resources, we need to have some idea of the temporal patterning of social relationships. At one extreme, if groups of bats are defined on the basis of roosting in a cave together and no bats change caves during the day, then there is no point in spending all day watching the bats in one cave. Instead, just enter the cave for long enough to identify the animals using it, move to the next cave and identify its inhabitants, and end the day's observation when all caves, or some predetermined proportion of them, have been sampled. At the other extreme, if fish are continually changing their associations, longer periods spent with any individual or group or in any area may be appropriate.

Clearly, to study social structure over any time scale means that we need data over that scale. Thus, if the important social time scales are unknown, it makes sense to arrange the data collection so that a range of scales can be examined. Thus, if we collect data for 6 hours per day on days 1, 2, 3, 11, 12, 13 June; 1, 2, 3, 11, 12, 13 July; and 1, 2, 3, 11, 12 August, we can examine scales of up to 5 hours, 1 to 3 days, about 10 days, 30 days, and 60 days within a total of roughly 100 hours of observation, a quite diverse set of spans. It is also important to consider the timing of important activities, such as breeding periods, that may affect interactions, associations, and social structure.

3.7.3: To Follow? What to Follow: Individuals or Groups? Another important decision when planning studies of social structure is who to watch and for how long. An initial choice is whether to follow or survey. In a *survey*, an individual or group is first encountered and then observed, and then the researcher, or her eyes or binoculars, move on to another individual or group. In a *follow*, the researcher's attention stays with an individual or group. The survey-or-follow decision is sometimes trivial, but at other times, it is more challenging.

If one is studying a small, captive, low-energy population in an open habitat, then perhaps all individuals and each action can be seen. In such cases, it is possible and optimal to follow everyone, recording interactions and/or noting changes in association and groups. At the other extreme are large, active populations most of whose habitat is invisible. Dolphins of the open ocean are an example. We cannot consistently follow either individuals or groups or observe interactions. We are constrained to survey individuals as they are encountered and to note associations or groups. In intermediate situations, a range of factors comes into play when choosing an observational strategy.

To record interactions (events), we must follow, at least for short periods, because, by definition, no interaction is visible in an instantaneous survey. If associations (states) are the measures of choice or necessity, however, then surveys will be generally more efficient if the rate at which individuals change associations is less than the rate at which new groups of animals can be surveyed (Whitehead 2004). For example, suppose group composition changes about once an hour; then, in terms of producing a model of social structure from records of group membership, it is more efficient to leave each group after noting its membership as long as another group can probably be found within 1 hour.

If the decision is made to follow, then should individuals or groups be the subject? Obviously, the more animals on which data can be collected simultaneously, the more powerful is the analysis. If the group is small and interactions are infrequent and easily seen, then group follows in which all interactions among all individuals are recorded are optimal. More normally, however, if interactions are the social measures of choice, then these will be difficult to record systematically for a whole group (Altmann 1974). Thus, if interactions are being recorded, then usually they should be between a focal individual that is being followed and others. In particular circumstances (such as parent–offspring or courting pair), dyadic follows may be appropriate, with interactions between the focal pair and between the focal pair and others all being recorded. Although recording all interactions within a group is usually impracticable (Altmann 1974), however, it is often possible to record all associations, especially if using the "gambit of the group" so that the group itself is used to define associations. In such cases, group follows will provide more information per unit time.

It is sometimes possible, and profitable, to make hybrid follows. For instance, while tracking a group of sperm whales (*Physeter macrocephalus*) for periods of days, we may carry out focal-individual follows of individuals during the 8 to 10 minutes that animals spend at the surface between dives (Whitehead 2004). Similarly, when following a large group of ungulates, one can make surveys of subgroups as they are encountered.

Frequently, "ideal" protocols for collecting social data are modified for reasons that are strategic (e.g., the desire simultaneously to collect data for another goal, such as population analysis) or tactical (such as weather).

3.7.4: Choosing Subjects. Although the random or systematic selection of experimental subjects is a cornerstone of statistical methodology (e.g., Sokal & Rohlf 1994, p. 393), in social analysis it is not so crucial. Clearly, if there are sets of individuals with very different social behavior, we need data on all of them, but it does not matter much if we gather relatively more data on some than on others. Hypothesis tests are not very frequent in social analysis, and those that are performed are usually framed in terms of the behavior of the individuals that have been sampled (e.g., "within the sampled population of individuals, males form larger groups than females"), in which we assume that we have obtained a random sample of the behavior of the sampled individuals, not that the sampled individuals are a random selection from the entire population.

Thus, when choosing the cave in which to identify roosting bats or which member of a captive population to begin a focal individual follow, we could use a random numbers table, but we could also use other criteria. Caves or focal individuals or classes of animals (Section 3.6) could be chosen in rotation, or a special focus could be placed on those that are deemed particularly interesting (perhaps mothers or caves with high bat densities). In more difficult research settings, subjects of surveys and follows are often chosen haphazardly, such as "the first group we come across." This is usually acceptable, even if it means that individuals with a home range near the research base are sampled more often than those who live at a greater distance.

Subject choice, however, could affect some social measures. For instance, if groups are surveyed as they are observed and large groups are more prominent than small ones, then an estimate of mean group size calculated from the data will be biased upward.

Another decision faces those who follow groups: When the group splits, which part should be followed? A rule such as "follow the largest of the daughter groups" will have no important effect on measures of dyadic relationship but would bias group size estimates. It is possible randomly to choose the daughter group that will be followed or, even better but impracticable in some circumstances, randomly pick a key individual when the group is first encountered and then, when a split occurs, follow whichever daughter group contains the individual.

3.7.5: Sampling Protocols. Altmann (1974) and others (e.g., Martin & Bateson 2007, pp. 48–61; Lehner 1998, pp. 195–210; Mann 1999) list a number of sampling protocols, such as "ad libitum," "focal-animal," "all-event," "predominant activity," "point," "scan," "1:0," or "sequence." These have different advantages, disadvantages, and recommended uses. In most formulations, the relative merits of the sampling protocols are confounded with follow protocols (discussed earlier). Here and in Table 3.4, I adapt the standard terminology and recommendations for the collection of social data, indexing choices by the follow protocol (surveys, individual follows, or group follows) and type of social measure being collected (interactions, associations or groups):

Surveys, recording associations or groups. Here the sampling is an instantaneous scan and usually 1:0 (a dyad is or is not associated, or are members or not members of the same group). Sometimes, however, individual or dyadic behavioral state data are collected. These could be ordinal or continuous data or categorical data with several states. They can be used to produce associations in subsequent analysis. For instance, locations can give nearest-neighbor data, whereas behavioral state and movement measures allow synchronicity to be assessed. In the simplest and most common format, the members of a group are noted during each survey of each group.

Individual follows, recording interactions. Ideally, each interaction involving the focal animal is recorded together with the time, type, and identity of the interactant. This is sometimes called "all-event" sampling. Simplifications include omitting the time information but recording the order of interactions (this becomes "sequence sampling" under some definitions), simply counting all interactions between each dyad ("sociometric matrix"), or recording whether there was an interaction between the focal animal and each other individual ("1:0 sampling"). I recommend that time, interaction type, and interactant identity be recorded if possible. Some measure of effort is required for most analyses, usually the time or number of sampling periods (Section 3.9) spent observing each individual.

Individual follows, recording associations or groups. There are several possible sampling protocols for association measures during individual follows. The associates of the focal individual can be recorded at regularly spaced instants (e.g., "who was the focal individual associated with at 12:05?," also called "point sampling"), during regular intervals (e.g., "who was the focal individual associated with between 12:05 and 12:10?," a form of "1:0 sampling"), or when they changed (e.g., "A became associated with the focal individual at 12:07," a form of "all-event sampling"). As long as the intervals used are not greatly longer than the rates of disassociation, it probably makes little difference to the results which of these is used. As with the survey protocols, individual data (e.g., identities and positions relative to the focal animal) can be recorded for all nearby individuals using any of these methods and then used later to produce one or more association measures with the focal animal (such as nearest neighbor).

Group follows, recording interactions. This will only be possible in rare cases in which the group is small and easily viewed and interaction rates are low, but in such cases, "event sampling" (in which events are the interaction types) is appropriate and efficient.

Group follows, recording associations or groups. The possibilities are similar to those available for association measures during individual follows listed previously, but with a few additional options. The simplest, and probably most frequently used, sampling protocol is simply to list group membership at regularly spaced sampling points ("point sampling") or whenever it changes ("event sampling"). Alternatives are to note associations within the group, if association is defined other than by membership of the followed group, either directly (e.g., subgroup membership) or by recording individual data such as position and behavioral state that can be used to derive associations later. Sometimes, these data will only be collected for a subset of the group.

It is often possible and desirable to combine sampling protocols. For instance, interactions and associations can both be recorded during individual follows. Sometimes, the data collected can be used to derive two or more association measures, such as "behavioral coordination" and "within *x* body lengths." Finally, the maligned ad libitum sampling method (basically field notes) should be used to record unusual but important behavior, such as fighting or mating, whether or not the focal individual is a participant and whether or not the behavior occurs during a survey or follow (Altmann 1974).

3.7.6: Effects of Observers. It is important both ethically and scientifically to minimize the effects of observation. Disturbed animals may form larger or smaller groups or increase or decrease their rates of association or dis-association (e.g., Foster & Rahs 1983; Kinnaird & O'Brien 1996), often showing antipredator-type behavior (Frid & Dill 2002). Martin and Bateson (2007, pp. 17–18) and Lehner (1998, p. 210) discuss causes and remedies for observer effects. Similarly, effects on behavior caused by individual identification (Section 3.5) and the collection of data for classifying animals (Section 3.6) should be minimized.

3.7.7: Nonobservational Data. It has been tacitly assumed throughout this section that interaction or association data are collected through visual observation, which may be real time or by analysis of video or still images. There are, however, other sensory modes. Interactions can sometimes be heard. Associations can be measured in a large range of ways (Table 3.2), including the co-occurrence of natural (e.g., DNA analysis of discarded body tissue or feces) or artificial (e.g., PIT tag) individual markers, as well as through the products of nonsocial analyses (such as the overlap of ranges).

3.8 Data Formats

Database and spreadsheet software are almost essential for storing social data (Section 2.9), but what format should be used? In this section, I recommend formats that either allow relatively simple manipulation in spreadsheet programs such as Excel or are suitable for my software package, SOCPROG. Other specialized software packages, such as UCINET and MatMan (Section 2.9), assume some processing of the raw data into similarity or dissimilarity matrices (Section 2.5). The preferred format may depend on whether interactions, associations, or groups are recorded directly or are derived from other recorded measures. Finally, I suggest a format (the SOCPROG format) for entering supplemental data, such as age or sex, that directly or indirectly can be used to allocate individuals to classes (Section 3.6).

First, a few preliminaries. I suggest that dates and times be combined in one field using the database or spreadsheet date–time format. Second, changes and ambiguity in field (column) formats can cause problems in

Table 3.5 Example of Interaction Data with Two Asymmetric (X and Y) and One Symmetric (M) Interaction Types Coded in Linear Mode for Situations in Which All individuals Are Observable

Date and time	Type of interaction	Actor/recipient	Interaction no.	ID
12/9/89 9:01	X	0	1	A1
12/9/89 9:01	X	1	1	A9
12/9/89 9:22	X	0	2	A14
12/9/89 9:22	X	1	2	A15
12/9/89 12:10	Y	0	3	B8
12/9/89 12:10	Y	1	3	A11
12/9/89 12:17	X	0	4	A13
12/9/89 12:17	X	1	4	A20
12/9/89 15:32	M	1	5	A4
12/9/89 15:32	M	1	5	A7
12/9/89 15:44	X	0	6	B12
12/9/89 15:44	X	1	6	A17
12/10/89 9:09	Y	0	11	A19
12/10/89 9:09	Y	1	11	A1
12/10/89 9:40	M	1	12	A9
12/10/89 9:40	M	1	12	A14

Asymmetric association data can be coded similarly.

analyses within Excel (and probably other spreadsheet software) as well as when the data are exported into other programs (such as SOCPROG). Therefore, I suggest that one not identify some individuals (or behavior types or classes) by numbers (such as "1453") and others alphanumerically (such as "53c"); one should just use numbers or alphanumeric codes throughout each field, whichever is more appropriate.

Data are usually stored so that rows represent observations, and columns (fields) the circumstances of the observation, what was observed, and who was observed. In SOCPROG, the final field gives the identities of the observed individuals, and I stay with this convention in the examples given later. I distinguish three ways of coding social data:

1. *Linear mode* (e.g., Table 3.5), in which each row corresponds to one observation of one individual. This is a SOCPROG format.
2. *Dyadic mode* (e.g., Table 3.6), in which each row corresponds to an observation of an association or interaction of a dyad. Thus, there are two identity fields representing the two identities in the dyad. This is a particularly useful format for asymmetric interactions or associations. Occasionally, the two identities may be of the same individual, as when an individual grooms itself (e.g., Table 3.6) or the presence of

Table 3.6 Example of Asymmetric Interaction or Association Data Coded in Dyadic Mode for Situations in Which All Individuals Are Observable[a]

Date and time	Groomer ID	Groomee ID
1/1/00 9:49	131	202
1/1/00 14:54	142	155
1/1/00 15:41	176	194
1/2/00 9:11	194	202
1/2/00 9:41[a]	100	100
1/2/00 10:09	6	162
1/3/00 10:35	100	188
1/3/00 11:03	196	202
1/3/00 14:32	6	162
1/3/00 17:40	155	89
1/4/00 7:16	196	202
1/4/00 13:17	131	3
1/4/00 16:15	155	89
1/5/00 6:00	51	89
1/5/00 15:57	162	100
1/5/00 17:55	131	3
1/11/00 7:19	188	127
1/11/00 10:09	89	45
1/12/00 7:14	89	45
1/12/00 9:01	162	100

[a] Including one case in which the interaction is of an animal with itself.

a noninteracting or nonassociating individual needs to be noted (see later discussion). This is a useful format for processing in Excel or other spreadsheet programs, for instance, by using "pivot tables" to produce counts of interactions or associations.

3. *Group mode* (e.g., Table 3.7) in which observations of one, two or more than two individuals are represented on each row, and one field gives all the identities of the individuals observed in the group. This is a SOCPROG format, and is compact, using less computer space and memory than individual or dyadic mode to store the same data.

Dyadic and group mode data can always be converted to linear mode data,[2] and linear mode data can usually be converted to dyadic mode. Linear and dyadic mode data cannot necessarily be converted into group mode, however, because linear and dyadic mode data are not necessarily symmetric and transitive, a requirement for group mode data (if A and

2. SOCPROG can convert group mode data into linear mode.

Table 3.7 Example of Coding Symmetric Association Data with One Association Type Such as Group Membership, Coded in Group Mode, Collected from Surveys

Date and time	Associating IDs
1/1/00 9:49	6 13 20
1/1/00 14:54	15
1/1/00 15:41	17 19
1/2/00 9:11	20
1/2/00 9:11	10 18
1/2/00 10:09	6 16
1/3/00 10:35	5 10 18
1/3/00 11:03	20
1/3/00 14:32	6 10 16 20

B interact/associate and B and C interact/associate, this does not necessarily imply that A and C interact/associate).

3.8.1: Coding Interaction Data. Coding interaction data is not always straightforward. The simplest case occurs when the whole population and all their interactions of certain types are observed and all interaction types are symmetric, so that there is no ordering to the interaction. Then, we can use dyadic or group mode data storage, with each row representing an interaction. The fields will usually contain date/time, type of interaction if more than one is observed and the identities of the interactants, as in Tables 3.6 and 3.8. Additional fields may contain information such as place and intensity of the interaction. Interactions involving three or more individuals can also be coded in this way in group mode simply by having more than two individuals recorded in the ID field, or in dyadic mode by having, for three individuals, three rows representing all three dyadic interactions.

If interactions are not symmetric, as in grooming (A may groom B without B grooming A) or fight outcomes, then we can use group mode data storage, but the first individual listed is considered the actor and the second (or perhaps all of the others if more than two individuals are listed) the receivers. Dyadic mode is particularly well suited for this situation (e.g., Table 3.6), or it may be best to use linear mode to code the data. One field represents the interaction type and another whether an individual is the actor or recipient, and a third distinguishes the different interactions, as in Table 3.5, which codes a mixture of symmetric and asymmetric interaction types. For each observed asymmetric interaction one individual has a "0" in the actor/recipient field and another individual a "1." For symmetric interactions, both individuals have a

Table 3.8 Example of Interaction Data with Five Symmetric Interaction Types Coded in Group Mode for Situations in Which All Individuals Are Observable

Date and time	Interaction type	Interactant IDs
1/1/00 9:49	A	13 20
1/1/00 14:54	A	14 15
1/1/00 15:41	A	17 19
1/2/00 9:11	F	19 20
1/2/00 9:41	F	10 18
1/2/00 10:09	A	6 16
1/3/00 10:35	A	10 18
1/3/00 11:03	D	19 20
1/3/00 14:32	A	6 16
1/3/00 17:40	A	15 8
1/4/00 7:16	D	19 20
1/4/00 13:17	A	13 3
1/4/00 16:15	A	15 8
1/5/00 6:00	A	5 8
1/5/00 15:57	B	16 10
1/5/00 17:55	F	13 3
1/11/00 7:19	C	18 12
1/11/00 10:09	A	8 4
1/12/00 7:14	C	8 4
1/12/00 9:01	B	16 10

"1." This format can also be extended for triadic interactions or those including more than three individuals.

An additional, but very important, consideration in most circumstances is the coding of control data. In addition to recording the observed interactions, we need to know for which individuals we could have recorded interactions had they taken place, so that, for instance, rates of interaction per unit time can be calculated for each dyad. Thus, in cases such as focal animal or group follows, where not all members of the population are being observed all the time, effort data must be coded in some way. This can be done by including, at least once per sampling period, data on "null interactions" that simply note the animals that could have interacted. For focal group follows, this can be achieved in group mode format. With individual follows, however, linear or dyadic mode will usually be needed as if A is being followed and B and C are also being observed such that they could have interacted with A; then we need to record the possibility of AB and AC interactions, but not BC ones. Tables 3.9 to 3.11 give, respectively, examples of group, dyadic, and linear mode interaction records with control data.

3.8.2: Coding Direct Association Data. When symmetric associations are recorded directly, the data can be coded using dyadic mode, as in Table

Table 3.9 Example of Interaction Data with Two Symmetric Interaction Types Coded in Group Mode, Including "Null" Interaction Effort Data ("0"), as Might Be Obtained from a Focal Group Follow

Date and time	Interaction type	Interactant IDs
1/1/00 9:40	0	13 18 20
1/1/00 9:45	0	13 18 20
1/1/00 9:49	A	13 20
1/1/00 9:50	0	13 14 20
1/1/00 9:55	0	13 14
1/1/00 14:45	0	14 15 18 20
1/1/00 14:50	0	14 15 18 20
1/1/00 14:54	A	14 15
1/1/00 14:55	0	15 18 20
1/1/00 14:58	B	15 20
1/1/00 15:00	0	15 18 20

Table 3.10 Example of Asymmetric Interaction Data Coded in Dyadic Mode, Including "Null" Data—Individuals Who Could Have Groomed but Did Not—as Might Be Obtained from a Focal Follow of #131

Time	Grooming?	Groomer ID	Groomee ID
06:22:00	Yes	131	202
06:27:00	Yes	131	155
06:27:00	No	176	131
06:32:00	No	176	131
06:37:00	Yes	131	176
06:37:00	No	131	162
06:42:00	No	131	188
06:47:00	No	131	202
06:52:00	Yes	6	131
06:52:00	Yes	131	89
06:57:00	No	131	202
06:57:00	No	131	89

3.6, or group mode, as in Table 3.7, in which each row corresponds to animals that are associated with each other. If more than two identifications are noted in group mode, then each is assumed to have been associated with all the others. The coding is basically the same whether the data come from surveys, group follows, or individual follows and whatever sampling protocol is used. It is important, however, that all individuals observed within a sampling period are noted, with an individual that was not associated with any other being indicated by a single identification in a row in group mode (as in Table 3.7), or, in dyadic mode, as an association of an animal with itself. With asymmetric associations (such as nearest-neighbor measures), then linear or dyadic mode coding is required, as in Tables 3.5, 3.6, 3.10, and 3.11.

Table 3.11 Example of Interaction Data with Two Asymmetric (X and Y) Interaction Types plus "Null" Data ("0") Coded in Linear Mode as Might Result from Focal Individual Follows

Date and time	Type of interaction	Actor/Recipient	Interaction#	ID
12/9/89 8:55	0	1	1	A1
12/9/89 8:55	0	1	1	A9
12/9/89 8:55	0	1	2	A1
12/9/89 8:55	0	1	2	B10
12/9/89 9:00	0	1	3	A1
12/9/89 9:00	0	1	3	A9
12/9/89 9:00	0	1	4	A1
12/9/89 9:00	0	1	4	B10
12/9/89 9:01	X	0	5	A1
12/9/89 9:01	X	1	5	A9
12/9/89 9:05	0	1	6	A1
12/9/89 9:05	0	1	6	A9
12/9/89 9:20	0	1	7	A14
12/9/89 9:20	0	1	7	A15
12/9/89 9:25	0	1	8	A14
12/9/89 9:25	0	1	8	A15
12/9/89 9:30	Y	0	9	A14
12/9/89 9:30	Y	1	9	A15
12/9/89 9:35	0	1	10	A14
12/9/89 9:35	0	1	10	A15
12/9/89 9:40	0	1	11	A14
12/9/89 9:45	0	1	12	A14

A1 and A14 are the focal individuals.

3.8.3: Coding Indirect Association Data. Sometimes, associations are not recorded directly but are inferred later. In this case, data are generally recorded in linear mode. Fields may include date/time, position (one, two, or possibly three dimensional), heading, or behavior (events or states). Then one can derive association measures such as "nearest neighbor," "dived within 30 seconds of one another," or "within three body lengths and heading the same direction ($\pm 30°$)." SOCPROG can usually produce such association measures reasonably easily. Table 3.12 shows an example of such data.

3.8.4: Coding Group Data. Group data are the simplest to code, usually in group mode, as in Table 3.7, in which each row corresponds to a group. Once again, it is important that single animals are entered, as a row containing just one ID.

3.8.5: Coding Supplemental Data. For social analysis, in addition to data on interactions or associations, we generally use attributes of individuals to place them into classes (Section 3.6). Individual attributes can be used to calculate nonsocial relationship measures, such as age differences

Table 3.12 Example of Data Coded so That Associations Can Be Derived Later

Date and time	Position on branch	Branch no.	ID
12/9/89 9:01	12	1	A1
12/9/89 9:01	27	1	A9
12/9/89 9:01	31	1	A14
12/9/89 9:01	37	1	A15
12/9/89 12:10	5	2	B8
12/9/89 12:10	15	2	A11
12/9/89 12:17	6	3	A13
12/9/89 12:17	9	3	A20
12/9/89 15:17	21	3	A4
12/9/89 15:17	25	3	A7
12/9/89 15:17	29	3	B12
12/9/89 16:40	31	4	A17
12/10/89 9:09	19	5	A19
12/10/89 9:09	25	5	A1
12/10/89 9:09	31	5	A9
12/10/89 9:09	50	5	A14

The identities of birds perching on surveyed branches are recorded together with the position on the branch (in centimeters from the trunk of the tree). Associations such as "nearest neighbor" and "on same branch and within 15 cm" can be calculated from these data.

Table 3.13 Example of Supplemental Data That Assign Individuals to Classes (e.g., Sex), Can Be Used to Derive Classes (e.g., from Age, One Can Derive Age Classes), or Can Be Used to Produce Nonsocial Relationship Measures (Such as Haplotype Similarity)

ID	Sex	Age (yr)	Haplotype
1	M	15.5	A
2	M	2.7	H
3	F	5.8	H
4	M	14.5	G
5	M	20.8	F
6	F	9.7	A
7	F	7.4	F
8	F	24.6	G
9	M	6.1	H
10	F	17.2	A
11	M	11.7	A
12	M	17.7	F
13	F	11.7	A
14	M	4	B
15	M	15.7	C
16	F	0.3	A

or genetic relatedness (Section 4.2). Class allocations, or data used to produce them, can also be stored in spreadsheet or database format. Table 3.13 illustrates the format used by SOCPROG and UCINET. The first column (field) is a list of identification names or numbers, and the subsequent columns (fields) give information such as sex, age, or haplotype for the corresponding individual. If a spreadsheet program such as Excel is being used, supplemental data can be stored in the same file as the social data but using separate worksheets. When we have both social data and supplemental data, it is perhaps even more useful to use a relational database such as Access with linked relationships, which makes it simple to both change and view aspects of the data.

3.9 Sampling Periods

Time has two important roles in social analyses. First, because the temporal patterning of interactions and associations is one of the key attributes of a relationship (Hinde 1976), temporal methods should play a key role in the analysis (Sections 4.6 and 5.5). At a more basic level, for almost all statistical techniques, we need to define a "sampling period"—the temporal units of the analysis. Thus, in each sampling period, we may produce counts of dyadic interactions or abstract whether a dyad was associated or not.

There are a number of considerations in selecting a suitable sampling period, including natural breaks in the sampling scheme, the rate of data collection, and independence of neighboring periods. For instance, if sampling is only carried out in darkness or only in daylight, then a sampling period of a day has a natural break and may be appropriate. A sampling period so short that there are few data collected within it (e.g., few interactions observed) is rarely useful. At the other extreme, valuable information is lost if the sampling period is so long that, for instance, almost all individuals in the population have associated with each other during each period. If association data are being collected and associations in consecutive intervals are almost always identical, then the sampling period is probably too short, whereas if they are almost uncorrelated, then the sampling period may be too long. For most analyses, statistical independence between neighboring sampling periods is neither needed nor desirable because we are interested in how dyadic relationships change over a range of time scales (Section 4.6). However, there are exceptions. Some permutation tests (Bejder et al. 1998) and estimates of the power of social analyses (Section 3.11), assume independent sampling periods, independent in the sense that the data from neighboring sampling periods are no more alike than those from well-separated periods. Thus, for different analytical techniques, it may be appropriate to divide the data into sampling periods of different durations. Occasionally, it may be useful to use sampling periods defined by a measure other than time, such as the field study or survey.