of it. You must take into account whether or not you have chosen the right subjects, where you will work on them and when. You must consider how much your presence will affect them and be careful about how much you project human intentions and emotions on to them. You must consider their particular perceptual abilities and you should be very sensitive to the ethical issues raised by working on animals and the different issues raised by working on humans. You should also consider any pressures arising from the way in which you are funded.

3

Getting started

The steps involved in studying behaviour

Recipes for conducting research are rarely followed precisely and most scientists build their ways of investigation in periods of apprenticeship when they model themselves on the behaviour of more experienced colleagues. In considering the steps listed below you should be aware that many programmes of research enter the sequence at different points. In general, though, studying behaviour involves a number of inter-related processes in roughly the sequence in which we have listed them. We have described these steps in outline. Lehner (1996) and Hailman and Strier (2006) provide much fuller accounts of research methodology, although we depart from their schemes – particularly in the emphasis we have placed on our first five steps.

1. Ask a question

Before any scientific problem is investigated, some sort of question will have been formulated. The question may initially be a broad one, stemming from simple curiosity about a species or a general class of behaviour, such as 'What does this animal do?' Such a question is not a hypothesis.

The value of broad description arising from sheer curiosity should not be under-estimated. Alternatively, it may be possible at an early stage to formulate a much more specific question based on existing knowledge and theory, such as 'Do big males of this species acquire more mates than small males?' This is tacitly a hypothesis. Not surprisingly, research questions tend to become more specific as more is discovered about a particular issue.

The particular choice of question (or questions) may be influenced by a variety of factors, including previous knowledge, interests and observations made in the course of other research and the priorities of the group in which you work. Sometimes the impetus for a study stems from little more than a hunch or from a wish to see what an animal will do next.

2. Make preliminary observations

A period of preliminary observation is generally invaluable in deciding what measurements to make and should be regarded as a crucial part of any study. Jumping straight in and collecting 'hard data' from the very beginning is rarely the best way to proceed.

3. Identify the behavioural variables that need to be measured

The form of the research and the variables that are to be measured should then be chosen so as to provide the best account of what you have observed. Definitions of behavioural categories should be clear, comprehensive and unambiguous. Write down the definitions before starting to collect data.

4. Choose suitable recording methods for measuring these behavioural variables

No observer can record behaviour without selecting some features from the stream of events and ignoring others. This selection inevitably reflects the questions you asked at the beginning of the study. It simply is not possible to record everything that happens, because any stream of behaviour could, in principle, be described in an enormous number of different ways. Practise the recording methods, assessing the reliability and validity of each category. Drop categories that are clearly unreliable and irrelevant. Measure inter- and intra-observer reliability at the beginning and end of data collection (see Chapter 7). Be prepared to add new categories and to redefine categories in the light of preliminary observations and pilot measurements.

5. Collect and analyse the data

Use the same measurement procedures throughout. Attempt to plan in advance how much data you will need to collect in order to obtain a clear conclusion (see Chapter 8). Once embarked on the collection of data, some people find it difficult to stop. Use the appropriate statistical tools for analysing the data. In some studies, this may be the point when you present your findings to a wider audience. You may also want to move to a more experimental phase, in which case the next step is crucial.

6. Formulate precise hypotheses

A clear hypothesis invites a direct test, but remember that hypotheses may be tested by observing natural variation in a population as well as by performing experiments. The study may have started with a particular hypothesis or it might have arisen from a more open-ended phase of the work covered by steps 1 to 5 above.

Formulating hypotheses is a *creative* process, requiring imagination as well as some knowledge of the issues involved. Others may already have made suggestions to explain the phenomenon of interest. It is then often a good idea to think about the weaknesses of their proposals as a guide to formulating a new hypothesis. More generally, do not reinvent the wheel: find out what others have done and speculated upon in the area and build on their work, without being so restricted by what they claim that you neglect your own ideas.

It is not possible to give definitive advice on how to formulate good hypotheses, any more than advice can be given on how to write good literature or paint good pictures. Sometimes, though, it is worth considering hypotheses that have been particularly successful in adjacent areas of research to see if they can be adapted to provide a good explanation in the area under investigation. The aim should be to find the best explanations. That means not just any old explanation compatible with the data obtained so far, but one that unifies with simplicity, giving a common account of superficially diverse phenomena.

In general, the larger the number of plausible competing hypotheses that are formulated the better, particularly when they make different predictions. The danger with having only one hypothesis is that it may be more difficult to abandon it when its predictions are not supported by the evidence. Data may always be explained in more than one way. That said, it is often a good plan to ask yourself which of the competing hypotheses would provide the most unifying explanation for the data.

7. Make predictions from the hypotheses

Making the transition from thinking about a problem and formulating hypotheses to tackling it empirically is often one of the most difficult parts of research. A clear hypothesis should, by a process of straightforward reasoning, give rise to one or more specific predictions that can be tested empirically. The more specific the predictions are, the easier it usually is to distinguish empirically between competing hypotheses, and thereby to reduce the number of different ways in which the results could be explained. A failed prediction may not necessarily mean that the hypothesis was wrong; the method of testing it may have been at fault.

8. Design the tests

Even if steps 2 to 4 have already have been passed through, it may be necessary to consider them again. The variables that are to be measured should then be chosen so as to provide the best test of the different predictions made by competing hypotheses. Experiments can be greatly improved by careful design and thoughtful use of control groups. Good design allows fewer subjects to be used overall when several treatments are combined in a single experiment (see Chapter 8). Moreover, if the treatments combine to influence the outcome, such interactions between them will be uncovered. The numbers of subjects used must, however, not be so small that the study will reveal no clear conclusions. Advice from colleagues expert in the design of experiments is almost always helpful.

9. Run tests of your hypotheses

Use the same measurement procedures throughout and try, if possible, to collect data 'blind' so that you do not unconsciously select data that fit your hypotheses. Stop collecting data when you have reached a pre-determined threshold that enables you to provide clear and reliable answers to your questions. Some inexperienced observers will stop when they get a statistically significant result, forgetting that with small sample sizes flukes are more likely to arise. On the other hand, do not go on collecting data simply because it is possible to do so. Once collected, make sure that your data are properly labelled, dated and include crucial information about the conditions and by whom the data were collected.

10. Analyse the results of your tests

Prepare the data in spreadsheets so that they are easily inspected and made available for subsequent statistical analysis. You may need to combine measures and guard against some common mistakes when treating data points as though they were independent. Employ the appropriate statistical tools, both for presenting and exploring the data, and for testing the hypotheses. Carry out *exploratory* data analysis to obtain the maximum amount of information from the data and to discover unexpected results that generate new questions. Do not sacrifice clarity for complexity by using unnecessarily complicated statistics. Use *confirmatory* analysis to test hypotheses. Distinguish between testing existing hypotheses and generating new ones (see Chapter 9).

11. Consider alternative interpretations of the evidence

Do not draw more conclusions than the data support, but do try to formulate a list of questions and ideas suggested by the data that could form the basis of future research. Be prepared to consider a range of alternatives. Try to come up with the best explanation that would, if correct, fit with background knowledge. The aim is to unify and offer coherence where it might have been lacking before.

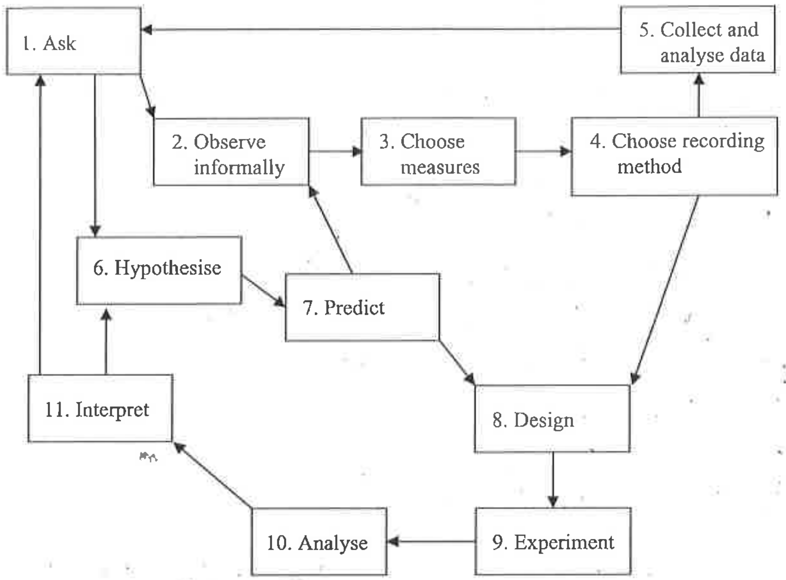

Figure 3.1 Summary of the steps involved in research. Some research will start at step 6 but would need to incorporate steps 2–4 before step 8. The progressive character of research may involve returning from step 11 to step 6 but in some cases may involve returning to step 1.

At this stage you may be ready to present your work to a wider audience orally or in written form. Always maintain scrupulous honesty. Submit your written work to peer-reviewed journals. When preparing your work for publication, pay attention to aspects of your work that may bear on matters of public interest.

As in any other area of science, measuring behaviour in order to discriminate between one set of hypotheses will inevitably produce results that in turn generate new hypotheses. In this sense, scientific research has a cyclical nature. It may be especially helpful to develop explanatory models that simplify complex phenomena, help you to understand them and to plan new work.

The steps involved in studying behaviour are summarised in Fig. 3.1. The successful scientist is likely to be one who can combine a purposeful approach to tackling an initial set of questions with the ability to recognise

and respond imaginatively to new questions that arise during the course of research. A study is unlikely to be fruitful if it remains open-ended and never focuses on specific issues. Conversely, if one problem is pursued in a rigid and inflexible way to the exclusion of all else then potentially important new ideas and observations may be missed.

Preliminary observation

Quantitative recording of behaviour should be preceded by a period of informal observation, aimed at understanding and describing both the subjects and the behaviour you intend to measure. Preliminary observation is important for two reasons: first, because it provides the raw material for formulating questions and hypotheses; and second, because choosing the right measures and recording methods requires familiarity with the subjects and their behaviour. Preliminary observation is especially important if the problems or animals are new to you.

Even in relatively open-ended studies, what you observe will reflect, in part, existing knowledge and theories, as well as your own preconceptions. Measurement of behaviour depends greatly on your being familiar with the animals you will observe, both from direct experience of watching them and by reading the literature about their biology and behaviour. A period of preliminary observation also provides a valuable opportunity for sharpening up your ideas and practising recording methods.

We cannot over-state the importance of simply watching before starting to measure systematically. This is particularly important when you are working with a species that has not been studied much previously. Beginners may falter early in a study because they rush to obtain 'hard data' and do not allow sufficient time to watch, think and frame interesting questions. Even an experienced observer needs to spend time on preliminary observation.

As a practical guideline, we suggest that you should always *plan* to exclude data obtained during the first few recording sessions from your final analysis. Otherwise it can be tempting to use all the data that you obtained, even though data from early recording sessions may be unreliable or not comparable to later data because of 'observer drift' or deliberate changes in measurement procedures. After a period of trial recording sessions, in which behavioural categories and measurement techniques have been tried out, preliminary data should be analysed. It is at this stage that methods should be modified if necessary.

Describing behaviour

Behaviour can be described in a number of different ways. The simplest distinction is between describing behaviour in terms of its structure or consequences.

The **structure** is the appearance, physical form or temporal patterning of the behaviour. The behaviour is described in terms of the subject's posture and movements.

The **consequences** are the effects of the subject's behaviour on the environment, on other individuals, or on itself. In this case, behaviour may be described without reference to how the effects are achieved. Categories such as 'obtain food' or 'escape from predator' are described in terms of their consequences, and may be scored irrespective of the actual pattern of body movements used.

For example, 'turn on light' is a description in terms of consequences, while 'press switch down using index finger' is a structural description. Similarly, 'run tip of bill along primary feather of wing' is a structural description, while 'preen' is a description by consequence.

Describing behaviour by its structure can sometimes generate unnecessary detail and place demands on your ability to make subtle discriminations between complex patterns of movement. Description by consequence is often a more powerful and economical approach, and has the additional advantage that the consequences can sometimes be recorded using automatic devices.

It is not uncommon for behaviour to be described in terms of presumed consequences or causes that later turn out to be wrong. Because of this danger it is best to use neutral terms for labelling behaviour patterns, rather than labels that falsely imply knowledge of the animal's internal state or the biological function of the behaviour pattern. For example, if a category of vocalisation is named *distress call* (rather than given a neutral

label such as *peep*), then you might be tempted to include vocalisations that did not meet the stated criteria for the category, but which were emitted when the animal was apparently distressed.

A third form of description is in terms of the individual's **spatial relation** to features of the environment or to other individuals. In this case, the subject's position or orientation relative to something (or someone) is the salient feature. In other words, the emphasis is not on what the subject is doing, but where or with whom. For example, 'approach' or 'leave' might be defined in terms of changes in the spatial relation between two individuals.

Choosing categories

Behaviour consists of a continuous stream of movements and events. Before it can be measured, this stream must be divided up into discrete units or categories. In some cases behaviour appears to be composed of natural units of clearly distinguishable, relatively stereotyped behaviour patterns such as pecks or grunts, and the division process will partly be dictated by the behaviour itself. To a large extent, though, the suitability of a category depends on the question being asked rather than on some inherent feature of the behaviour. Observational categories must reflect some sort of implicit theory and do not have an existence of their own, independent of the observer. It is therefore difficult to give specific advice on what sorts of categories to choose, although we can offer some general guidelines.

- Enough categories should be used to describe the behaviour in sufficient detail to answer the questions and, preferably, to provide some additional background information.
- Each category should be precisely defined and should summarise as much relevant information as possible about the behaviour.
- Categories should generally be independent of one another; that is, two or more categories should not be merely different ways of measuring the same thing.
- Categories should generally be homogeneous: that is, all acts included within the same category should share the same properties.

Inexperienced observers often err on the side of trying to record too much. A given stream of behaviour could potentially be described in an almost limitless number of ways, depending on the questions being asked, so it is essential to be selective. It is certainly best to drop categories that are clearly irrelevant, or which seem inconsistent and difficult to measure reliably. The chances are that the fewer categories used, the more reliably each will be measured.

Bear in mind, though, that you will improve with experience, so data from later recording sessions may be reliable even if data from early sessions are not. Furthermore, it may be better to record too much initially, rather than too little. Redundant or unreliable categories can always be discarded or pooled at the analysis stage. It is also wise to collect supplementary information that might in the future provide useful background or raise new questions. However, collecting a wide range of measures or supplementary information should not be allowed to detract from careful measurement of the important things.

The extent to which the definitions of individual categories are specific rather than general will depend on the nature of the problem. Questions and hypotheses tend initially to be rather broad and then narrow down as more is discovered about a particular problem. The more clearly and precisely the initial question has been formulated, the more obvious it will be what to measure.

When choosing categories it can sometimes be helpful to have descriptions of the main types of behaviour pattern that typify the species. In some cases this information is available in the form of an 'ethogram', which is ostensibly a catalogue of descriptions of the discrete, species-typical behaviour patterns that form the basic behavioural repertoire of the species. Unfortunately, published ethograms vary enormously in the number of behavioural categories included and the detail with which these are described, and ethograms are unavailable for many commonly studied laboratory subjects. Moreover, ethograms are frequently of limited use because not all members of a species behave in the same 'species-typical' way. On the contrary, individuals of the same species, even when of the same sex and age, can behave in quite different ways.

Defining categories

Each category of behaviour to be measured should be clearly, comprehensively and unambiguously defined, using criteria that can be easily understood by other observers. More important still, the criteria used to define a category should unambiguously distinguish it from other categories, particularly those it resembles most closely. A detailed and complete definition of each category and the associated recording method should be written down *before* the data used in the final analysis are collected. Two types of definitions are used:

Operational definitions specify the physical operations that are required by the researcher to make the measurement; these are most commonly used when measuring the consequences of behaviour.

Ostensive definitions involve giving an example of the case to which the category applies using diagrams and written descriptions, for example when an individual plays with an object. These definitions are most often used in the direct observation of behaviour. An ostensive definition of an action should enable another observer to recognise the same pattern of behaviour.

The period of preliminary observation provides an opportunity to develop the precise criteria used to define each category. A completely satisfactory and unambiguous definition of a category can rarely be formulated without having watched the behaviour for some time. Preliminary definitions are often unable to deal with unforeseen ambiguous examples of the behaviour that crop up during preliminary observations, and must be therefore modified in the light of experience.

Clearly, all the data for a particular category that are used in the final analysis must be strictly comparable. Thus, data obtained before the final definition of a category was formulated must be discarded. Developing a set of precise and unambiguous category definitions can be a slow process.

Writing down precise definitions of categories at the beginning of the study is essential to prevent definitions and criteria from 'drifting' during the course of the study (see Chapter 7 on factors affecting reliability). Written definitions should be sufficiently precise and detailed to enable another observer to record the same things in the same way.

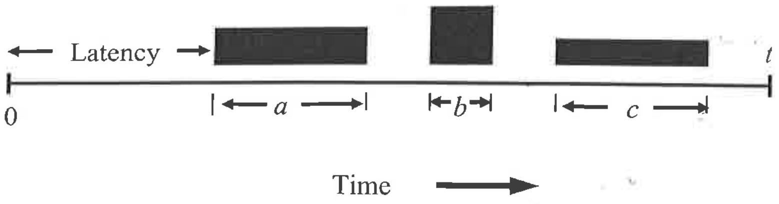

Figure 3.2 The meaning of latency, frequency, duration and intensity. The black rectangles represent three successive occurrences of a behaviour pattern during an observation period of length t units of time. Latency is the time from the beginning of the observation to the first occurrence of the behaviour. Frequency is the total number of occurrences divided by the total observation time ($3/t$). The total duration of the behaviour is $a + b + c$ units of time and the mean duration is the total duration divided by three. Intensity is the amplitude of the behaviour represented by the height of the rectangles.

Types of measure

Behavioural observations most commonly yield four basic types of measure (see Fig. 3.2).

Latency (measured in units of time: e.g. s, min or h) is *the time from some specified event* (for example, the beginning of the recording session or the presentation of a stimulus) *to the onset of the first occurrence of the behaviour*. For example, if a bird does not approach a novel object until 6 min have elapsed, the latency to approach is 6 min. If, as is normally the case, the period of observation is limited and each individual is tested more than once, then the behaviour pattern may not occur at all during some tests. A related measure to latency is the lag between one event and another – say between one animal performing an act and another performing the same act.

Frequency (measured in reciprocal units of time; e.g. s^{-1}, min^{-1} or h^{-1}) is *the number of occurrences of the behaviour pattern per unit time*. Frequency is a measure of the *rate* of occurrence. For example, if a rat presses a lever 60 times during a 30-min recording session, the frequency of lever pressing is 2 min^{-1}.

An alternative usage, which is perhaps more common in the behavioural literature and in statistics, is when 'frequency' refers to the *total number of occurrences*. However, this usage is uninformative and potentially misleading, unless the total time for which the behaviour was watched is also specified. For example, to state that the 'frequency' of a behaviour was 60 is meaningless: did it happen 60 times in two minutes; one hour; a day . . .? Most statements about total numbers of occurrences could equally well refer to rates of occurrences, since a total number of occurrences can always be expressed as a rate, assuming the length of the observation period is known. To avoid any confusion, the *total number of occurrences* should be explicitly referred to as such. Expressing frequencies in the way we suggest (number per unit time) removes any possible ambiguity.

Duration (measured in units of time: e.g. s, min or h) is *the length of time for which a single occurrence of the behaviour pattern lasts*. For example, if a kitten starts suckling and stops 5 min later, the duration of that period of suckling was 5 min.

'Duration' is also used in at least two other senses in the behavioural literature. The first is when 'duration' (or 'total duration') refers to the *total* length of time for which all occurrences of the behaviour lasted over some specified period, usually the whole observation session. A total duration is, of course, meaningless unless the total time for which the behaviour was watched is also specified. For example, to state that the 'total duration' of a behaviour pattern was 16 min says nothing: was it 16 min out of 20 min; 30 min; an hour; a day . . .? To avoid any ambiguity, we recommend that a total duration should be expressed as the total duration over the specified period of observation (for example, '9 min per 30 min') and should be explicitly referred to as **total duration**.

Alternatively, a total duration can be expressed as a proportion (or percentage) of the observation period, in which case it should be explicitly referred to as the **proportion** (or percentage) of time spent performing the behaviour. For example, if a kitten spent a total of 10 min suckling during a 30-min observation session, then the proportion of time spent suckling was $10/30 = 0.33$. Note that a proportion or percentage of time is a dimensionless index with no units of measurement.

Expressing a duration as a proportion or percentage of total time omits the potentially important information about the total time for which the behaviour was watched. For example, the interpretation placed on the statement that the proportion of time spent sleeping by a subject was 0.10 must depend on whether this figure was based on, say, a 24-h period of observation as opposed to a 30-min observation.

'Duration' (or 'mean duration') is also used to refer to the *mean* length of a single occurrence of the behaviour pattern, measured in units of time (e.g. s, min or h). This is obtained by recording the duration of each occurrence of the behaviour pattern and calculating the mean of these durations. To avoid any possible ambiguity, we suggest that this measure should be referred to explicitly as a **mean duration**.

Mean duration can also be calculated by dividing the total duration of the behaviour pattern by the total number of occurrences. This has the advantage that the duration of each occurrence need not be recorded separately; you could, for example, use a cumulative stopwatch to record the total duration and a counter to record the total number of occurrences.

As an illustration of these various measures, suppose that a mammalian mother and offspring are observed for 60 min, during which suckling occurred five times, the individual periods of suckling lasting 3 min, 10 min, 1 min, 1 min and 1 min, respectively. According to our suggested definitions, the *durations* of suckling were 3,10,1,1 and 1 min; the *total duration* of suckling was 16 min per 60 min; the *proportion of time* spent suckling was 0.27 (= 16/60); and the *mean duration* of suckling was 3.2 min (= 16/5).

Frequency and duration, which are the measures most commonly used for describing behaviour, can give different and complementary pictures. For example, how often two monkeys groom each other (frequency) tells us something different about the nature of their social relationship from how long they spend doing it (duration). Frequency and duration measures of the same behaviour are not always highly correlated, so it is probably wise to record both.

Intensity. In general, categories are best defined in such a way that the behaviour is simply scored according to whether or not it has occurred or for how long it has occurred, rather than making assessments of intensity or amplitude. Intensity, unlike latency, frequency and duration, has no universal definition. Nonetheless, it may be helpful or even essential to make judgements about the intensity or amplitude of a behaviour pattern.

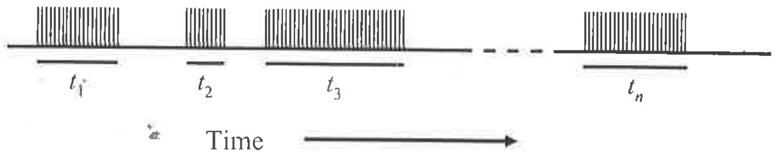

Figure 3.3 The meaning of local rate, which provides another measure of the intensity of behaviour. An activity such as eating may be composed of discrete component acts such as ingesting a morsel of food, indicated here by vertical lines. The local rate is given by the total number of occurrences of the component act (number of food morsels eaten during the observation period) divided by the total duration of the activity ($t_1 + t_2 + t_3 + \ldots + t_n$)

In some cases the consequences of the behaviour can be measured in terms of some physical quantity related to the behaviour, such as the weight of food eaten, the volume of water drunk, the number of prey captured or the distance travelled. It may be useful to measure the sound intensity of a vocalisation, the amplitude of a limb movement or the height of a jump; or to estimate the intensity of a facial expression or the aggressiveness of a social interaction. Intensity can sometimes be measured according to the presence or absence of certain components of the act, which may be present at high intensity but absent at low intensity.

A simple and informative index of intensity is **local rate**, defined as the number of component acts per unit time spent performing the activity. For example, suppose that an activity – eating – is composed of discrete, component acts – the ingestion of individual items of food. The local rate of eating would, in this case, be given by the number of items ingested per unit time spent eating (see Fig. 3.3). Similarly, the intensity of walking might be measured by the number of strides per unit time spent walking, and the intensity of grooming by the number of face-stroking paw movements per unit time spent grooming. Local rate captures the speeded-up or hurried nature of intense behaviour: the more hurriedly the activity is performed, the higher its local rate.

Events and states

When choosing the type of measure to describe a behaviour pattern, it is helpful to distinguish between two fundamental types of behaviour pattern which lie at opposite ends of a continuum.

Events are behaviour patterns of relatively short duration, such as discrete body movements or vocalisations, which can be approximated as points in time. The salient feature of events is their *frequency* of occurrence. For example, the number of times a dog barks in one minute would be a measure of the frequency of a behavioural event.

States are behaviour patterns of relatively long duration, such as prolonged activities, body postures or proximity. The salient feature of states is their *duration* (mean or total duration, or the proportion of time spent performing the activity). For example, the total time a dog spends asleep over a 24-h period would be a measure of the total duration of a state. (Note that the term 'state' is also used in the behavioural literature to refer to a motivational state, such as hunger or thirst, so it is important not to confuse the two.)

The onset or termination of a behavioural state can itself be scored as an event and measured in terms of its frequency.

The different levels of measurement

Measurement means assigning numbers to observations according to specified rules. Four different levels of measurement are distinguished, in ascending order of strength of measurement:

Nominal. If observations are simply assigned to mutually exclusive, qualitative classes or categories, such as male/female or active sleep/quiet sleep/awake, then the variable is measured on a nominal (or categorical) scale. If only two outcomes are possible (e.g. yes/no, or male/female) the data are said to be binary.

Ordinal. If the observations can also be arranged along a scale according to some common property then the variable is measured on an ordinal (or ranking) scale. The number assigned to each measurement is its **rank**. For example, if in a given period of time individual *A* played more than *B* who played more than *C*, then *A* would be ranked highest on this measure, *B* next and *C* lowest.

Interval. If, in addition, scores can be placed on a scale such that the distance between two points on the scale is meaningful – i.e. the *difference* between two scores can be quantified – the variable is measured on an interval scale. The zero point and unit of measurement are arbitrary for an interval scale. A temperature measured in degrees Celsius is measured on an interval scale.

Ratio. The highest level of measurement is attained when the scale has all the properties of an interval scale but also has a true zero point. This is referred to as a **ratio** scale since, unlike an interval scale, the ratio of any two measurements is independent of the unit of measurement. Mass, length and time are measured on ratio scales. For example, the ratio of two periods of time is the same whether they are measured in seconds or days. True frequencies, durations and latencies are measured on ratio scales.

Assigning numbers to a category of behaviour does not necessarily mean that the behaviour is measured on an interval or ratio scale. For example, subjectively rating an individual's aggressiveness on a scale of 0 to 5 would not constitute measurement on a true interval scale, since there is no reason to assume that the difference between scores of, say, 1 and 2 is the same as the difference between scores of 4 and 5. Such scores can be ranked, but the differences between them are probably not meaningful and therefore the measurement would be on an ordinal scale. A ratio scale is simply not attainable in much behavioural work.

Summary

The sequence of operations required in most studies of behaviour is shown in Fig. 3.1. Attempts to test a specific hypothesis need not be involved at the outset of a study. Later in the sequence, however, hypotheses will be formulated and these should lead to direct tests. A period of preliminary observation is essential in any study. How the activities of interest are described and categorised will depend on the questions that are being asked. So too does the level of measurement.