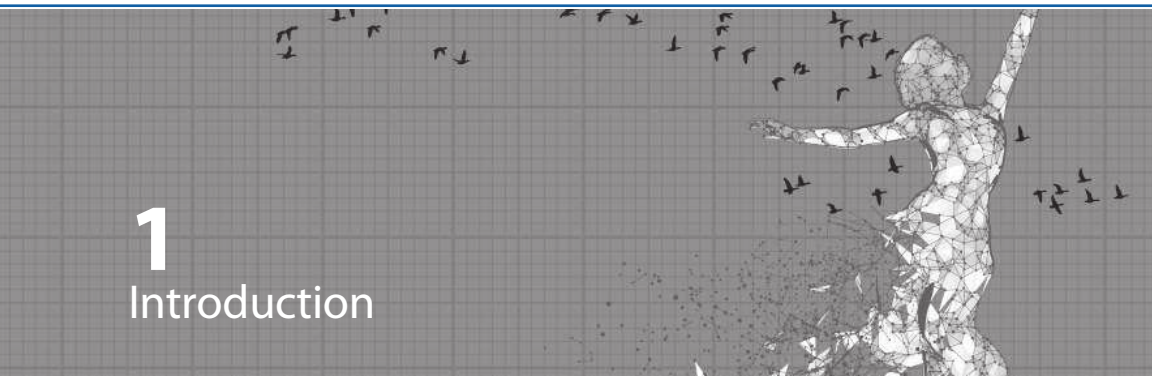

1 Introduction

1.1 Scope and Aims

As its title suggests, this book is about measuring behaviour. By behaviour, we mean the actions and reactions of whole organisms or groups of organisms. Examples of behaviour include the mating displays of bower birds in the jungle, the foraging patterns of laboratory mice and the productivity of humans working in an office. The organisms whose behaviour is measured are usually multicellular animals such as insects, birds and mammals, including our own species, *Homo sapiens*. However, the possession of a brain is not a prerequisite for demonstrating measurable behaviour: unicellular organisms such as bacteria and protozoa, and even plants, are increasingly targets for behavioural research [1]. The 'organisms' need not even be biological: they might be artificial intelligence (AI) systems embodied in robots or autonomous vehicles, or virtual agents behaving in the virtual environment of a computer [2]. This book is intended to be relevant to anyone interested in measuring behaviour in any of these diverse entities, which we refer to as 'subjects'.

Measuring behaviour accurately and reliably requires various problems to be solved. General principles underlying the measurement and analysis of behaviour apply whether the subjects are microbes, animals, people or machines. Many of the same problems and principles apply whether the behaviour being studied is occurring in the natural environment, the laboratory or a computer. Our aim is to describe these problems and principles and offer practical advice to anyone who wants to understand how behaviour is measured.

The main focus of this book is on the direct observation of behaviour. By direct observation, we mean the description and analysis of what subjects actually do in a specific situation. Direct observation does not imply that the behaviour must be directly observed in real time but rather that the variables recorded must relate directly to the actual performance of the behaviour in question. In the past, behaviour was generally measured

by human observers using low-tech methods such as check sheets, voice recorders or computer event recorders, and such methods are often still appropriate. Increasingly, however, researchers are using automated methods such as machine vision, data loggers attached to subjects and, for human subjects, smartphones.

Some behaviour patterns may leave semi-permanent evidence that can be measured as a proxy for the behaviour of interest. For example, the structure of a bird's nest provides a record of its construction decisions; bald patches of skin on a mouse can be evidence of over-grooming by cage mates; and a photograph of a meal may provide a reasonable proxy for human diet choice.

In non-human animals, the direct observation of behaviour or its immediate proxies is the only option available for measuring behaviour. For humans, however, surveys, interviews and questionnaires are commonly used to measure how subjects believe they behaved or might behave in given situations. While such methods involve the performance of behaviour – the subject has to make verbal or written responses – the behaviour that is directly observed (e.g. ticking a box on a questionnaire) is distinct from the actual performance of the behaviour of interest (i.e. the behaviour described in the questionnaire) [3]. Measurements based on self-reported memories, or beliefs about probable behaviour, may be justified on the grounds that they are less time consuming than direct observation and may correlate reasonably well with actual behaviour. Nonetheless, retrospective self-report is often very different from what people actually did, owing to a combination of inaccurate memory and biases in what people believe about how they behave, or what they choose to convey to researchers. For example, total calorie consumption is systematically underreported by obese human subjects [4]. Alcohol consumption is also systematically underreported, with retrospective reports becoming more inaccurate the more subjects drank on the day they were asked to recall [5]. Retrospective self-report measures therefore provide an indirect and often inaccurate means of measuring behaviour, though they are sometimes the only feasible means available.

Previous editions of this book were focused on measuring behaviour in non-human species. However, we wanted this new edition to be equally relevant to the direct observation of behaviour in humans. Modern technologies offer novel approaches to the direct observation of human

behaviour that circumvent traditional obstacles. For example, wearable fitness-tracking devices can measure actual physical activity, and smartphones can be used to ask subjects what they are doing in real time. Although the latter method relies on self-report, it is less susceptible to inaccuracies resulting from poor memory and also to biases resulting from what is socially or medically acceptable. For instance, admitting to drinking three units of alcohol that evening is still well within an advisory limit, whereas 21 units that week puts the subject over the current UK advisory limit and may be information the subject would choose not to share.

Measurement means recording phenomena in a systematic, reproducible way and expressing the results in numerical form. Measuring behaviour therefore implies a **quantitative** approach. Social scientists make a distinction between quantitative and **qualitative** research methods. To explain the difference, let us consider how quantitative and qualitative researchers might explore whether a group of chimpanzees is unusually aggressive. Both types of researcher would spend time watching the animals. The quantitative researcher would then define a set of aggressive behaviours (e.g. chases, threats, bites) and count the occurrence of these behaviours observed in a specified time period. They would use statistical methods to summarise the number of aggressive behaviours seen in the group and compare the results to comparable data collected from other groups to test the hypothesis of a difference in aggression. The qualitative researcher would describe verbally what they observed without the constraint of sticking to predefined categories. They would identify themes in aggressive behaviour (e.g. male–male aggression, inter-sexual conflict, abuse of infants) and use specific examples to illustrate them (e.g. Bobby bit Kathy on the arm when she attempted to grab his food). Based on this data, they would reach a subjective impression of the types of aggression displayed and the aggressiveness of the group. Thus, the distinction between qualitative and quantitative research comes down to the methods used to represent, summarise and analyse the data collected.

Qualitative approaches give a richer impression of the behaviour, but quantitative approaches make it clearer how conclusions follow from the data because they are more transparent and reproducible. Furthermore, quantitative data allows statistical comparisons to be made between groups and hypotheses to be formally tested. Worthwhile research on behaviour will require that at least some aspects of behaviour are measured, and we

would advocate a quantitative approach wherever possible. That said, all quantitative measurements are ultimately based on subjective judgements about what to ask and what to record. The art of measuring behaviour is learning to ask the right questions and to choose good behavioural metrics.

1.2 Why Measure Behaviour?

Behaviour is centrally important in many areas of biology, psychology, medicine and the social sciences. Within the biological sciences, behaviour sits at the nexus between neuroscience, cognitive science and ecology. Behaviour is the major output of the brain and is therefore central to unravelling the cognitive and neural mechanisms underlying it. Behaviour is also a major means by which individual organisms adapt to and affect their physical and biological environments, making it a major factor in ecology. Human behaviour is, of course, central to psychology and most fields of social science. Anthropology, communication, education, economics, human geography, law, linguistics, political science, public health, security and sociology all involve the description and understanding of human behaviour and its consequences. Many of the biggest problems facing the human species, including climate change, infectious disease pandemics, obesity, mental health, geopolitical conflict and terrorism, are reflections of human behaviour. The vital importance of measuring how people behave was highlighted by the COVID-19 pandemic. Accurate and reliable descriptions of the behaviour underlying such problems are critical to understanding them and implementing the changes necessary to address them. Given the central importance of behaviour in so many disciplines, researchers must understand how to measure it accurately and reliably, while policy makers must understand how to distinguish between good and bad measurements.

1.3 What is So Special About Measuring Behaviour?

The answer to this question lies in the characteristics of behaviour that make its measurement challenging. Behaviour has a temporal component – it unfolds over time. Therefore, it is rarely possible to measure behaviour in

a meaningful way by looking at a single snapshot in time; the methods used must accommodate the temporal component. Occasionally, there may be a simple proxy for a behaviour that *can* be measured at a single time point. For example, a single blood test for the level of cotinine (the main metabolite of nicotine) can reveal how much a person smokes on average. But it cannot reveal potentially important details about the pattern of smoking behaviour, such as the time of day at which the behaviour occurs or its social context.

Many behaviour types are not discrete, making it hard to determine exactly when they begin and end. Take play behaviour, for example. It can be relatively easy to recognise that play has occurred within a given time window but difficult to define its exact duration.

Behaviour patterns are often extremely complicated, making them difficult to represent with a single number or metric. Counting the stereotyped distress calls of a domestic chick is straightforward, whereas describing and quantifying the complex song of a nightingale in a meaningful way is considerably harder.

Behaviour often changes in response to stimuli from the environment. Moreover, different species have very different sensory systems, which means that stimuli that are undetectable to humans may be highly salient to other species. For instance, birds and bees can see in the ultraviolet (UV) part of the spectrum, while mice and rats can hear ultrasound. Even humans may respond to stimuli of which they are not consciously aware. For example, people's reactions to a movie may be heightened in the cinema by exposure to volatile organic compounds released by other viewers in response to humour or suspense [6].

A further important characteristic of behaviour is that it is highly variable. This variability occurs at different levels: there is variation in behaviour between individuals of the same species, depending on the individual's genes, sex, developmental history and so on; there is variation within individuals over time resulting from the effects of experience, maturation, developmental plasticity and senescence, among other things; and there is variation within individuals according to their current context. Behaviour varies according to physical variables, such as time of day and ambient temperature, and biological variables, such as hormonal state and social context. While some sources of variability will be obvious factors that can be accounted for when designing a study, others may be less apparent.

For example, the way in which a researcher handles a mouse when placing it in the test apparatus is known to affect the animal's anxiety levels even more than conventional laboratory manipulations of stress [7]. Similarly, the sex of the human experimenter present in the room is found to influence the behaviour of laboratory rodents in standardised tests of anxiety, probably via the volatile pheromones released by men and women [8]. These and other characteristics of behaviour make its measurement challenging. An awareness of the potential pitfalls is critical in designing and conducting a good study.

1.4 Steps Involved in Studying Behaviour

Those attempting to make systematic measurements of behaviour for the first time are often daunted by the apparent difficulty of the job facing them. How will they ever notice, let alone record accurately and systematically, everything that is happening? The truth is that measuring behaviour *is* a skill but not one that is especially difficult to master, given some basic knowledge and an awareness of the possible pitfalls.

Studying behaviour involves a number of inter-related processes that can be broken down into a series of steps. Although the steps listed below will apply in the majority of behavioural studies, the order in which they are taken will vary. Moreover, some steps may need to be repeated multiple times in the light of results obtained from preliminary observations or pilot studies. Box 1.1 gives an illustrative example of how the steps apply in an actual study. The steps involved in studying behaviour can be summarised as follows:

1. **Ask a question.** All scientific studies should start with a clear question.
2. **Formulate hypotheses.** Hypotheses are provisional and testable explanations for observed phenomena. A hypothesis can be thought of as a possible answer to the question. At least two alternative hypotheses should be formulated. Good hypotheses should give rise to testable predictions.
3. **Make predictions.** A clear hypothesis should, by a process of straightforward reasoning, give rise to one or more specific predictions that can be tested empirically. Developing specific predictions goes hand in hand with identifying which behavioural variables to measure.

Box 1.1 Example of the steps involved in a behavioural study exploring the relationship between self-reported food insecurity and measured snack food consumption in adult British women [9]

1. **Question.** Why is current food insecurity – defined as the limited or uncertain ability to acquire nutritionally adequate food – associated with obesity in women in affluent Western societies?
2. **Hypotheses.** Main hypothesis tested: food-insecure women are motivated to consume more calories than they require when the opportunity is available. Null hypothesis: the calorie consumption of food-insecure and food-secure women does not differ.
3. **Prediction.** When presented with a controlled opportunity to consume snack foods in the laboratory, food-insecure women (as defined by their responses to a standard questionnaire) will consume more calories than food-secure women.
4. **Behavioural metric.** The total calories in three snack foods (chocolate, crisps and popcorn) consumed by a participant in a fixed time period.
5. **Recording method.** Participants were given an opportunity to snack via the guise of a mock taste test in which they were asked to evaluate aspects of the palatability of three different snack foods. Plates of the three foods presented to the participants for evaluation were weighed on an electronic balance by the researchers before and after the test.
6. **Study design.** An observational study using naturally occurring variation in the food insecurity of an opportunity sample of participants recruited at Newcastle University. Recruitment of participants was time limited, but a minimum target number required was set prior to commencing data collection.
7. **Ethics.** The study was approved by the Faculty of Medical Sciences ethics committee of Newcastle University. All participants provided written informed consent and were debriefed after the study.
8. **Pilot study.** No formal pilot study was conducted in this case. Prior to collecting data, all the researchers involved in data collection rehearsed the protocol to ensure standardised procedures.
9. **Preregistration.** This study was not preregistered. It was therefore constrained to be published as an exploratory study.
10. **Data collection**. An agreed protocol was followed throughout, including a standard script for instructing participants. Researchers measured snack food consumption blind to the food-insecurity status of the participants. Data was collected from 84 women.

11. **Data analysis.** The hypothesis was tested using a general linear model, with total calories consumed as the outcome variable and food insecurity as a continuous predictor variable. As predicted, higher adult food insecurity was found to be associated with greater total calorie consumption in the mock taste test. Additional exploratory data analysis showed that greater self-reported childhood food insecurity reduced the positive effect of adult food insecurity on calorie consumption.
12. **Communication**. The study was initially written up independently by three undergraduate psychology students as a requirement for their degree course. Subsequently, their supervisors prepared a version of the study for publication in a peer-reviewed academic journal (*Appetite*). The raw data and an accompanying script for the data analysis were made publicly available via the Zenodo repository.

4. **Identify and define behavioural metrics.** Behavioural metrics must be defined clearly and unambiguously before starting to collect data. Metrics may need to be redefined or replaced in the light of preliminary observations and pilot measurements.
5. **Choose an appropriate recording method.** How will the metrics identified in step 4 actually be recorded to ensure accurate and reliable data? The method chosen will depend on the nature of the behaviour to be recorded and the technology available. In practice, steps 4 and 5 are inter-related, as the recording method often influences the choice and definition of metrics.
6. **Design the study.** This involves making choices about the type of study to be conducted (observing natural variation versus experimental manipulation), the selection of subjects (e.g. mice or humans), the allocation of subjects to experimental groups (sampling and randomisation), how many subjects are required and how data collection will be structured (cross-sectional versus longitudinal). Good design makes the difference between a study that allows hypotheses to be tested and one that is uninterpretable. It can also affect costs in terms of time, money and subject welfare, with implications for step 7.
7. **Ensure that the research is legally and ethically compliant.** Most behavioural research will require formal ethical permission to be obtained from

the relevant national or institutional regulatory bodies before it can proceed. A successful application for ethical approval will usually require that steps 1–6 have already been completed.

8. **Conduct pilot studies.** Jumping straight in and collecting 'hard data' from the very beginning is rarely the best way to proceed. Preliminary observations can be useful for formulating and sharpening hypotheses and refining behavioural metrics. Pilot studies allow researchers to practise experimental manipulations and recording methods and assess the reliability and validity of measurements.
9. **Preregister the study.** To reduce dubious research practices (see Chapter 2), it is good practice to preregister the hypotheses, predictions and methods for a study in a publicly accessible place before starting to collect data.
10. **Collect the data.** The same measurement procedures should be used throughout. If possible, data should be collected 'blind', so that measurements are not unconsciously selected or adjusted to fit the hypotheses. Data collection should stop when a predetermined threshold (identified in step 6) has been reached.
11. **Analyse and interpret the data.** The data collected should be thoroughly visualised and understood through the use of graphs and descriptive statistics before embarking on statistical hypothesis testing. Confirmatory data analysis is used to test preregistered hypotheses. Exploratory data analysis is used to obtain the maximum amount of information from the data and discover unexpected results that may generate new questions.
12. **Write up and communicate the research.** Research is wasted if it is not properly communicated. Where research has used sentient subjects, it is ethically dubious not to make the data and findings available because the subjects have given their time and in some cases their lives in the cause of research. Many research funders require data to be made publicly available.

In the following chapters we add flesh to the skeleton outlined above. Chapter 2 considers the quality of behavioural research and describes the practices necessary to ensure replicable findings. Chapter 3 distinguishes four logically distinct questions that can be asked about any behaviour and considers how a question is turned into testable hypotheses and predictions.

Chapter 4 discusses alternative study designs and the problem of how much data to collect. Chapter 5 describes the legal and ethical issues raised by behavioural research and the requirements before a study can proceed. Chapter 6 describes the various metrics available for quantifying behaviour and Chapter 7 the alternative rules for sampling and recording behaviour. Chapter 8 describes the technologies that assist data capture and processing. Chapter 9 discusses the problems and opportunities raised by the fact that subjects often occur in groups. Chapter 10 describes methods for ensuring the validity and reliability of behavioural measurements. Chapter 11 introduces how behavioural data is analysed. Finally, Chapter 12 considers pitfalls in the interpretation of findings and describes how research is communicated.

1.5 Summary

- Behaviour is the actions and reactions of an organism or group of organisms. Living organisms, robots and virtual agents all exhibit measurable forms of behaviour.
- Measuring behaviour involves assigning numbers to direct observations of behaviour using specified rules.
- Direct observation means collecting data that relates directly to the performance of the behaviour pattern in question.
- Measuring behaviour accurately and reliably is important because behaviour is central to answering many questions in the biological and social sciences.
- Measuring behaviour is challenging because behaviour has a temporal component, does not always occur in discrete bouts, is generally complicated, can be influenced by stimuli undetectable to humans, and varies both within and between individuals.
- Studying behaviour can be broken down into a series of steps that starts with asking a question and ends with communicating findings.