

2 An experiment

From vague thoughts to specific plans

This chapter covers the kinds of steps you need to make in developing a simple experimental design. You may have problems or questions of your own which you would like to turn into actual experiments. If so, good. Or you may have a topic given to you by someone else. Wherever the idea comes from there will be a need to move from something which is vague, woolly and ill-formulated to a specific plan of action.

Suppose one is interested in absent-mindedness; a topic which is familiar enough but which has, until recently, been largely neglected by the psychologist. The interested reader is referred to Baddeley (1981) who discusses work which he and others have done in this area as examples of studies of the cognitive psychology of everyday life.

The first step in designing an experiment is to try to get a reasonably explicit statement of the problem with which you are trying to deal. What is meant by absent-mindedness? It appears to be connected with not remembering to do something. I am reminded of a former colleague who would probably win prizes for absent-mindedness. While this showed itself in a variety of ways, his particular speciality lay in keys and cheque-books which were regularly mislaid: he would forget to bring keys with him when leaving the house or he would leave his cheque-book on the desk after using it, where it would disappear under a pile of papers. This was not a case of someone poor at either remembering or recall of material in the usual sense. Certainly he had an

An experiment

almost encyclopaedic knowledge of the published research in his area of specialism.

A possible design strategy might be to seek to relate these phenomena to some theory. Those with a Freudian leaning might see significance in the fact that it is keys and cheque-books which are forgotten. Seductive as such musings might be, Freud's concepts have proved notoriously difficult to translate into worthwhile experiments (whether this reveals shortcomings in Freudian theory or in the use of experimental approaches when seeking to understand humans is a separate issue which could well form the basis of another book). A more mundane approach, but probably more productive for the experimentalist, and perhaps more likely to lead to practical suggestions for dealing with the absent-mindedness, might be to analyse in more detail the circumstances in which the forgetting takes place. What might well be involved here is him not giving himself a cue to check something at a particular time. Is his difficulty that of checking that he has his key when leaving the house? And is this 'omitting to check' a common feature in cases of absent-mindedness?

Clearly one could move from this kind of anecdotal musing to design an experiment where one or more aspects of remembering to check something at a particular time are tested. This is the approach taken by Baddeley and his colleagues in a variety of tasks. These include having people return postcards to them at specified dates after they had been given them, and a simulation of the pill-taking regime of 'four times a day after meals' where volunteers had to press a button on a modified watch at four specified times each day.

However, there may be other facets of absent-mindedness exemplified by going to a café for a cup of tea and asking for a newspaper (or going to a newsagent and asking for a cup of coffee), as I have done on more than one occasion. This aspect was investigated by a group of students from Bexhill Sixth Form College in connection with a BBC television series 'Young Scientist of the Year' (further details are given in Baddeley, 1981). Their initial investigations of such 'slips of action' (other examples being answering the telephone by giving an address and trying to put on tights when wearing slippers) showed that there were quite consider-

Independent and dependent variables

able differences between individuals in the extent to which they reported such things as happening to them, and that the slips tended to occur when trying to perform a routine activity at the same time as doing something else.

The strategy of the Bexhill sixth-formers was to isolate two 'extreme' groups of people; one group who reported that they had no slips of this kind, and a second group who reported a lot of such slips. These two groups then participated in experimental situations where they had to perform two tasks at the same time.

This process of refining and clarifying the problem you are working on calls for a mixture of common sense and clear thinking. However, there is a quite substantial amount of jargon (or to put it more positively, technical language) associated with the design of experiments which you need to be familiar with – if only so that you can understand accounts of experiments, and communicate what you have done to others. Some of the important terms you need to know are covered in the rest of this chapter and the next one.

Independent and dependent variables

The terms 'independent variable' and 'dependent variable' were introduced on p. 4 when discussing what is meant by an experiment. However, it will do no harm to go through this again here.

What we are doing in a simple experiment is trying to observe the relationship between two variables. The variable which the experimenter manipulates is called the **independent variable** (this is often simply abbreviated as IV). The IV here is concerned with some aspect of absent-mindedness. In the design of the experiment we are going to have to be very clear as to what we mean by this – see the discussion below on 'operational definitions'.

You should note that the independent variable is being manipulated here by the way in which the experimenter selects the groups taking part in the experiment – the two extreme groups of 'high' and 'low' absent-mindedness. It would be difficult to directly manipulate the degree to which a particular individual is absent-minded. A similar, and very common, example occurs in experiments looking at gender differences where the experimenter selects

An experiment

female and male groups. This type of ‘manipulation by selection’ leads to complications in the interpretation of the results of the experiment, discussed in Chapter 8 (p. 132). There are, however, many situations where it is perfectly feasible to manipulate the independent variable directly by, for example, altering the type of material presented or in some other way changing the type of experience or situation for different groups.

The variable which is observed in order to see whether changes in the IV have any effect on it is known as the **dependent variable** (also commonly found in abbreviated form as DV). Here the DV is some aspect of performance of the tasks. Again this has to be very carefully defined – see the discussion below on ‘operational definitions’.

In psychological experiments the independent variable is very often a stimulus variable (e.g. the type of material to be learned, the brightness of a light, exposure-time of a word, etc.): that is, in general, the **input** to the person taking part. An important exception to this has already been noted. This is when the independent variable is associated with a feature of the people taking part in the experiment (e.g. ‘male’ or ‘female’). The dependent variable is almost always a response variable (time taken to make a response, strength of response, number of responses, etc.): that is, in general, the **output** from the persons taking part.

Qualitative and quantitative variables

Fairly obviously, a ‘variable’ is something which can vary. In other words, it can take on different values or levels. For instance, if the dependent variable is the number of errors made this might take on just about any whole number (integral) value. In an experiment where we are considering how problem-solving varies with age, the dependent variable would be age, and the values or levels of the variable used might be 4 years, 6 years, 8 years, 10 years, etc. Variables expressed in numbers in this way are referred to as quantitative variables. Dependent variables are almost always quantitative (if one includes in this category simple counts as to how many times a particular thing occurs) as this then opens the possibility of statistical analysis, which is regarded by conventional experimentalists as central to their approach.

The independent variable is, however, quite commonly qualitative rather than quantitative, as in the example of the use of gender as an independent variable. Here, any assignment of numbers such as 'female = 1' and 'male = 2' (which may be done, for example, in coding survey responses) is purely arbitrary. In our 'absent-mindedness' example the two values of the independent variable are effectively 'high absent-mindedness' and 'low absent-mindedness'. While they are derived from numerical values they are in fact treated qualitatively rather than quantitatively.

Experimental conditions

The values of the IV ('high absent-mindedness' and 'low absent-mindedness') are commonly called the **experimental conditions**. (You may also find them referred to as 'treatment conditions'.) In this experiment, and in almost all the other experiments considered in this book, we will deal with just two values of the IV (i.e., two experimental conditions). This is partly because the statistical techniques that will be covered can only deal with two conditions at a time. However, keeping within these limits, it is possible to answer a very large number of experimental problems. There are techniques for dealing with more than two conditions at a time, and these are covered in more advanced texts. Another possibility is to deal with more complicated experimental designs by considering the values of the independent variable two at a time. This kind of piecemeal approach is not recommended as it can throw away many of the advantages of using more complex designs.

Operational definitions

An **operational definition** is stated in terms of the steps or operations that have to be carried out in observing or measuring whatever it is that is being defined. Before we can make an idea for the experiment into an actual experiment that we can carry out, we must define our independent and dependent variables in this way.

In considering the independent variable and the specific experimental conditions (levels or values of the IV) we need to ask ourselves exactly what we mean. What aspect of 'absent-mindedness'

An experiment

are we focusing on, and what is 'low absent-mindedness' and 'high absent-mindedness'? Here the procedure (i.e. the steps or operations followed) was to ask for recording of the absent-minded 'slips' over a four-week period. Those reporting no instances over the period were assigned to the 'low absent-mindedness' group; those reporting more than eight to the 'high absent-mindedness' group.

As far as the dependent variable was concerned the exact nature of the two tasks had to be specified, together with exactly what was to be measured. The two tasks were backward counting in threes from a specified number and mirror-drawing. The latter, if you are not familiar with it, is a common laboratory task involving tracing around the outline of a star with the tracing hand only visible through a mirror – a surprisingly difficult and frustrating task. The measure in the first task was the number of items counted in a specified time and in the second task the time taken to complete the maze.

It is only when we have precisely defined the variables that the experiment can be carried out. Equally important, it is essential that we define the terms in this exact way so that some later worker, coming along and seeing our results, being interested in them or perhaps even disbelieving them, will then be able to set up an exact replica of our experiment in order to check on the result we have obtained.

This is called **replication** of the experiment. While it may not seem a particularly exciting or interesting task it is an important one and should probably take place much more frequently than it does currently in psychological and other research involving people. We may be building our disciplines on shaky foundations. For example, a very widely quoted and influential study 'Pygmalion in the Classroom' by Rosenthal (who has carried out much of the work on 'experimenter expectancy' effects referred to on p. 136) showed that teachers' expectations of children's performance, artificially manipulated in the experiment, brought that performance up to the expectations. Shipman (1988) points out that several attempts to replicate these 'findings' have been unsuccessful. Similarly, Baddeley notes with regret that the particular findings of the 'Bexhill' study have, to date, proved impossible to replicate. Perhaps we can leave further studies in this area as a challenge to the reader.

Subjects or participants?

The convention has been that those taking part in an experiment are referred to as **subjects**. The symbol S is used to indicate a subject (but it is sometimes used in statistics to stand for other things – so beware). The use of the term ‘subject’ is now criticized by some as indicating an inappropriate kind of relationship – being the ‘subject of investigation’ or ‘subject to the control of the experimenter’. It could be argued that this is in fact a fair description of the power relationships in the experimental situation, but I have to admit that once sensitized to such issues I now prefer the term **participant** – i.e. someone taking part in the experiment. Given that the British Psychological Society now uses this terminology in its ‘Ethical Principles for Conducting Research with Human Participants’ (BPS, 1993) I propose to make the switch to ‘participant’ in this text (sorry that it is longer than ‘subject’; and, unfortunately, the corresponding symbol P can also stand for other things in statistics).

Samples and populations

When carrying out an experiment the usual hope and intention is to try to find out something of relevance and applicability beyond the specific group of participants involved on a particular day in a particular place using particular equipment and materials. Mention has already been made of the necessary artificiality of much laboratory work which calls into question its relevance to real world, non-laboratory settings. In practice, judgements are made about wider applicability in terms of plausibility. Is this the kind of finding likely to have been different if the study had been carried out in Stockport or Stockholm rather than in Stevenage? In summer as against winter; or using flash cards instead of a tachistoscope? Such questions can be addressed directly by seeking to replicate the findings and establishing how robust they are.

However, a particular form of reasoning is commonly used in experimentation and depends on the notion of samples and populations. Specifically, if we can show that in carrying out an experiment we are dealing with a sample which is representative of a certain

An experiment

known population, then it is possible for us to generalize with a degree of confidence from the specific sample to the population which it represents. Although this kind of reasoning is not limited to making generalizations about people, i.e. from the sample of participants in the experiment to a population which they represent, it is this aspect which is most common and which illustrates the principle most clearly.

Suppose we are interested in carrying out experiments with children between eight and ten years old. Because virtually all children of these ages attend school it is not too difficult, given the necessary permissions, to obtain a list of such children within a particular Local Education Authority. This set of children might constitute the population for your studies. Note that unless you have a way of tracking down children not on the schools' books – e.g. those educated at home, or in private education outside the area, etc. – the population you are dealing with changes from all the children in the area to all those attending schools in the area. Even then there are grey areas; do you include all schools – opted out, private and special?

For a particular experiment you would select a representative sample of, say, sixty children from this population. Probably the best, and in principle (though not in practice) simplest way of doing this is through selection of a **random sample**. That means using some scheme which guarantees that each child in the population has an equal chance of appearing in the sample. Appendix 1 shows how this could be done. Such random sampling permits the generalizing of any findings about the sample tested to the population from which they are drawn and is central to much of the statistical reasoning developed later in the book.

It must be admitted, however, that in practice the use of random sampling from a known population to establish a set of participants is quite rare. Student experiments are commonly carried out on an incestuous basis where members of the class participate in each other's experiments or manage to persuade or cajole other students to take part. As the main purpose here is for the experimenter to develop skills and understanding in carrying out experiments, this shortcoming may not be too serious. Even in research carried out

Three basic designs or how do participants fit into the experiment?

primarily to extend knowledge it is not unusual to depend on similar 'convenience' samples using whoever can be persuaded to take part. Does this matter? In one sense yes, as such experimenters are almost certainly using statistical techniques based on assumptions about representative sampling. In other senses the answer is probably no: the experimenter is in fact much more interested in establishing possible causal relationships than worrying about the generalizability of results.

Three basic designs or how do participants fit into the experiment?

Given that we have chosen our IV and DV and decided on the way in which they will be operationally defined in the experimental situation, some of the most important decisions remaining relate to the way in which participants are assigned to the different conditions.

These ways of assignment lead to three basic experimental designs. These are given a variety of labels but will be referred to here as the **independent samples design**, the **matched pairs design** and the **repeated measures design**.

1 The independent samples design

For this design a group of participants is obtained for the experiment as a whole, and then individuals are allocated randomly to one or other of the experimental conditions. The term 'independent samples' arises from this randomness of allocation. If we decide to use sixteen participants in all, they will be allocated randomly to the two conditions. It is usual to do this allocation with the further stipulation or constraint that there are equal numbers in the two groups. This could be done using a coin with heads for condition A say, and tails for condition B, or, alternatively, using odds and evens from random number tables. We would continue allocating in this way until eight participants had been allocated to one of the two groups, when the remaining participants would be allocated to the other group.

An experiment

The allocation might be as follows:

Table 1 Allocation of participants in independent samples design

| condition A | condition B |
|-------------|-------------|
| P_1 | P_3 |
| P_2 | P_6 |
| P_4 | P_7 |
| P_5 | P_8 |
| P_9 | P_{10} |
| P_{12} | P_{11} |
| P_{13} | P_{15} |
| P_{14} | P_{16} |

Here P_1 stands for participant one, P_2 for participant two, etc.

In the experiment each participant provides a single score for purposes of analysis, i.e. the total number of participants is the same as the total number of scores.

2 The matched pairs design

In this design, participants are matched in pairs and the two members of each pair allocated randomly, one to each of the experimental conditions. Any experimenter having access to pairs of identical twins would be addicted to this design, but there are other ways in which pairs can be obtained. The matching can be performed in terms of a third variable for which the experimenter has good evidence that it is likely to affect scores on the dependent variable. Thus in a problem-solving task, it would be possible to match participants in terms of intelligence. If IQ test scores are available, pairs of participants could be selected who were closely equated in terms of their measured IQs. In the ‘absent-mindedness’ experiment it might, for instance, have been feasible to match participants in terms of their performance on some appropriate memory task.

The allocation might be as shown in Table 2. Here P_{11} stands for participant 1 in the first matched pair and P_{12} stands for participant 2 in the first matched pair. P_{21} stands for participant 1

Three basic designs or how do participants fit into the experiment?

Table 2 Allocation of participants in matched pairs design

| condition A | condition B |
|-------------|-------------|
| P_{12} | P_{11} |
| P_{21} | P_{22} |
| P_{31} | P_{32} |
| P_{42} | P_{41} |
| P_{51} | P_{52} |
| P_{62} | P_{61} |
| P_{71} | P_{72} |
| P_{82} | P_{81} |

in the second matched pair, and so on. The decision about whether participant 1 or participant 2 within each pair is allocated to condition A or to condition B is on a random basis. As with the independent samples design, each participant provides a single score for analysis, the total number of participants again being the same as the total number of scores.

3 The repeated measures design

In this design, a single participant appears under both of the experimental conditions. Thus for the same number of scores as the other two designs we only need half the number of participants.

Table 3 Allocation of participants in repeated measures design

| condition A | condition B |
|-------------|-------------|
| P_1 | P_1 |
| P_2 | P_2 |
| P_3 | P_3 |
| P_4 | P_4 |
| P_5 | P_5 |
| P_6 | P_6 |
| P_7 | P_7 |
| P_8 | P_8 |

Here P_1 stands for the same participant under both experimental

An experiment

conditions, P_2 stands for a second participant, who also appears under both conditions, and so on.

Although there are obviously no problems here in terms of the allocation of participants to the different experimental conditions there are special problems relating to the *order* in which each participant performs the two experimental conditions.

Order effects in repeated measures designs

The change in the DV produced by the change in IV is called an **experimental effect**. However, as we have previously discussed, there may well be changes in the DV produced by variables other than the IV – unless we are able to control the effects of these other variables. In a repeated measures design there may be some systematic effect associated with the order in which a participant is involved with the two conditions; i.e. an **order effect**. Thus it might be that in a particular situation there is a general practice effect (increased familiarity with the situation, ‘learning how to learn’, etc.) such that whatever is done second tends to get a higher score irrespective of any effect of the IV. Alternatively, it might be that there is a negative practice effect of a general kind (fatigue, boredom, etc.) such that whatever is done second tends to get the lower score.

In either case we cannot make unambiguous statements about the experimental effect, i.e. as to whether condition A or condition B produces the better results, because what we actually measure is the combination of the experimental effect and the order effect. Clearly, if each participant were to be tested with condition A first and then condition B afterwards the experimental effect is inextricably mixed up with any order effect (this is referred to as the effects of the two variables being **confounded**). Two methods are commonly used to try to sort this out.

Counterbalancing

In using counterbalancing some scheme is used so that half of the participants work under condition A first, and half work under condition B first. A simple version of this is shown in Table 4. Counterbalancing will only balance out an order effect in certain

Table 4 Counterbalancing participants

| condition A | condition B |
|----------------|----------------|
| P_1 (first) | P_1 (second) |
| P_2 (second) | P_2 (first) |
| P_3 (first) | P_3 (second) |
| P_4 (second) | P_4 (first) |
| P_5 (first) | P_5 (second) |
| P_6 (second) | P_6 (first) |
| P_7 (first) | P_7 (second) |
| P_8 (second) | P_8 (first) |

circumstances. It would, for example, do this where the order effect is adding a constant amount to the score of whatever comes second. Unfortunately, it is quite possible that the order effect is more complicated. There could be a large order effect when condition A comes first, and only a small order effect when condition B comes first. This kind of effect is called an **interaction** and if it occurs, counterbalancing will only partially balance out the order effect. The basic problem, however, is that we usually do not know the nature of any possible order effect and hence can't be sure that it has been adequately dealt with.

Randomization

An alternative to counterbalancing is randomization. This means we could decide by some random process, such as tossing a coin separately for each participant, whether he or she does condition A or condition B first. A fuller discussion of randomization and its relation to statistical inference follows in the next chapter. There are suggestions for methods of randomization in different situations in Appendix 1 (p. 149).

Individual versus group designs

The three basic designs discussed in the preceding section are all group designs. A group of participants are selected and, after their

An experiment

involvement in the experiment, comparisons are made between group scores on the dependent variable under the two experimental conditions. However, while in some areas of psychology (such as aspects of social psychology) we are interested in group behaviour in its own right, it is more usually the case that what we are really interested in is individual behaviour.

Why use groups then? The main reason is that humans are complex and even in well controlled experiments there is likely to be considerable variability in their response. Often so much so that it is very difficult to discern any effects of the independent variable in the face of what amounts to a large amount of random variability of individual response. Using both groups and statistical analysis helps in sorting this out. But it is at the expense of finding things out about the average participant, rather than any one individual.

There is an influential approach within experimental psychology pioneered by B. F. Skinner which rejects such group designs. It depends on demonstrating effects within the individual subject as changes from a steady 'baseline'. Skinner argues that, using his techniques, it should be possible to exert such a degree of control over extraneous variables that the effect of changes in the independent variable becomes so clear and obvious that statistical analysis is unnecessary. A clear exposition of this alternative methodology is given in Sidman (1960). It is particularly influential in some areas of applied psychology, for example in behaviour modification studies carried out by clinical or educational psychologists.

Figure 1 illustrates an example. The first step in this type of study is to establish a **baseline**. This is a steady state of responding over several periods or sessions which must be established prior to any intervention by the experimenter. When this has been achieved, the experimenter changes the experimental conditions and measures responding over a further set of sessions. In a simple version of this kind of design, the experimenter then returns to the initial baseline condition and again measures responding for several sessions.

The results shown in Figure 1 provide a convincing demonstration of the effect of the IV on the DV (rate of responding) because there is a substantial and stable change from a stable baseline, to which responding returns after the intervention. You might reasonably ask: what does one do if the picture is not as clear-cut as this?

Individual versus group designs

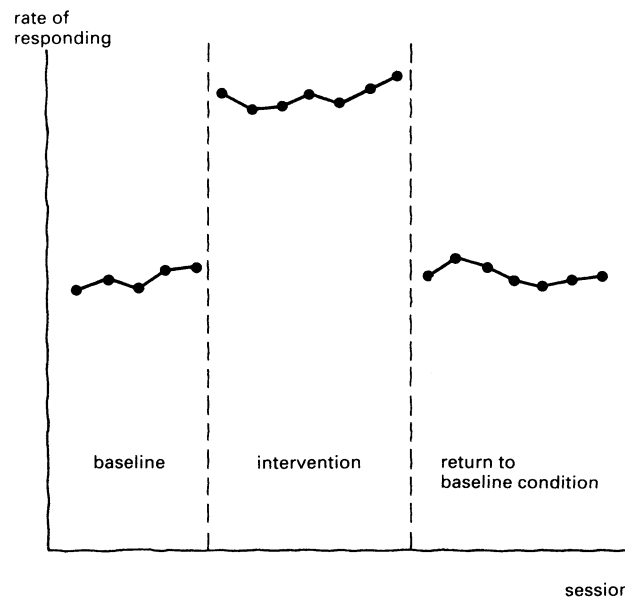


Figure 1 Example of simple baseline design

In this situation, Skinner puts the onus on you as experimenter to do a better experiment. You may need to control other variables more effectively so that the baseline becomes stable. Or slightly modify the intervention so that its effect becomes more clear-cut. Basically, the effects of the IV should be so unambiguous that you can simply 'eyeball' the data without need for statistical analysis.

This simple design (sometimes called an ABA design – referring to the baseline/intervention/and return to baseline phases respectively – has its problems. If used in an applied setting, where the intervention is usually seeking to bring about some desired change or improvement in behaviour, it is unfortunate and possibly unethical that the participant ends up where he or she started! So-called multiple-baseline designs and other more complex designs avoid such problems.

An experiment

Are you ready to start experimenting?

No. The discussion so far has covered some of the issues involved in designing an experiment. However, a basic rule of experimentation is that *you do not start unless and until you know how you are going to analyse the data you will obtain*. The next chapter provides a general introduction to the principles of statistical inference which underlie the analyses you might make. Following chapters provide a range of things that you might do with different kinds of data. Chapter 8 returns to matters of design.