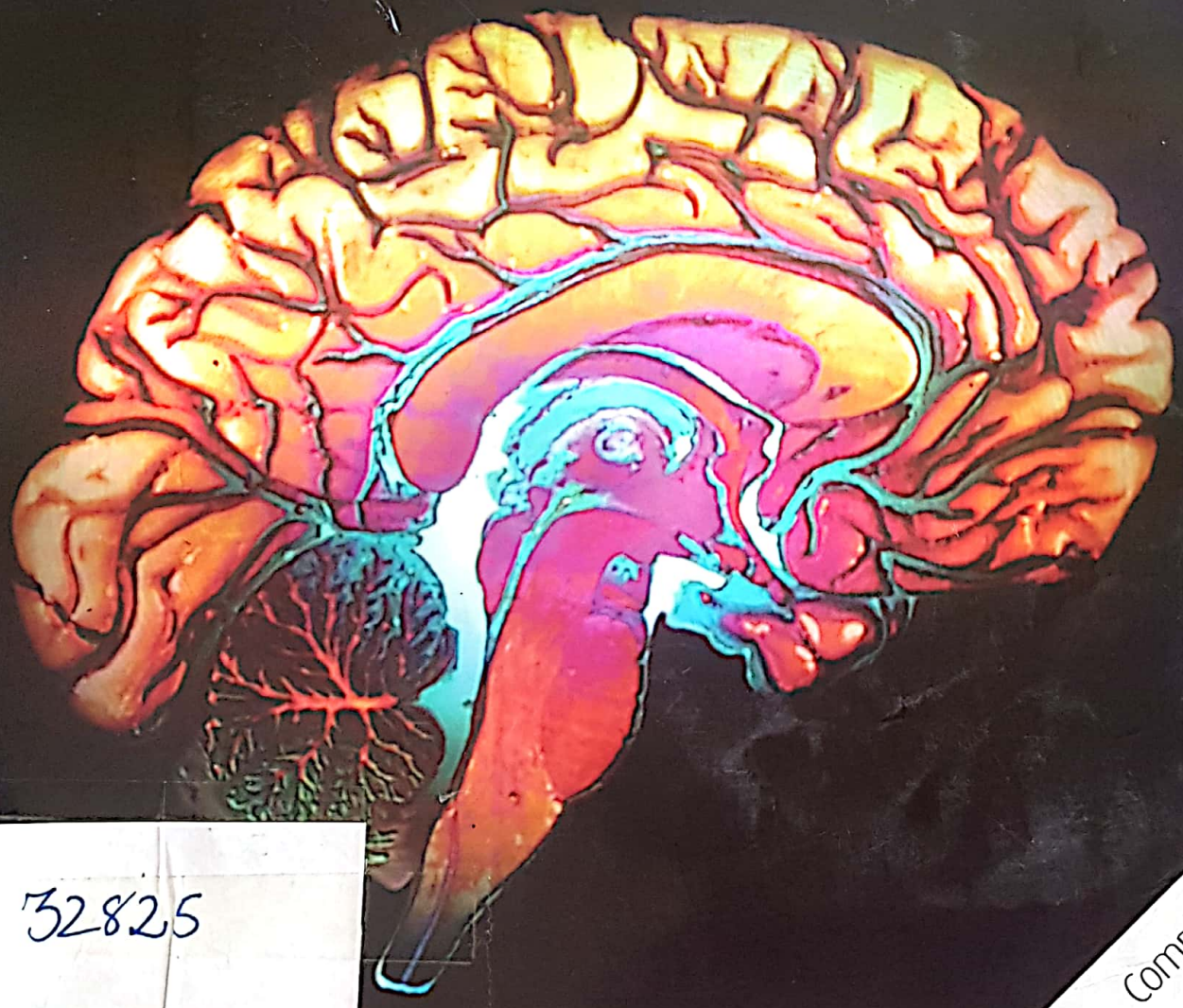
CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE International Examinations
Endorsed for full syllabus coverage

Julia Russell, Fiona Lintern,
Lizzie Gauntlett and Jamie Davies

Cambridge International AS and A Level Psychology

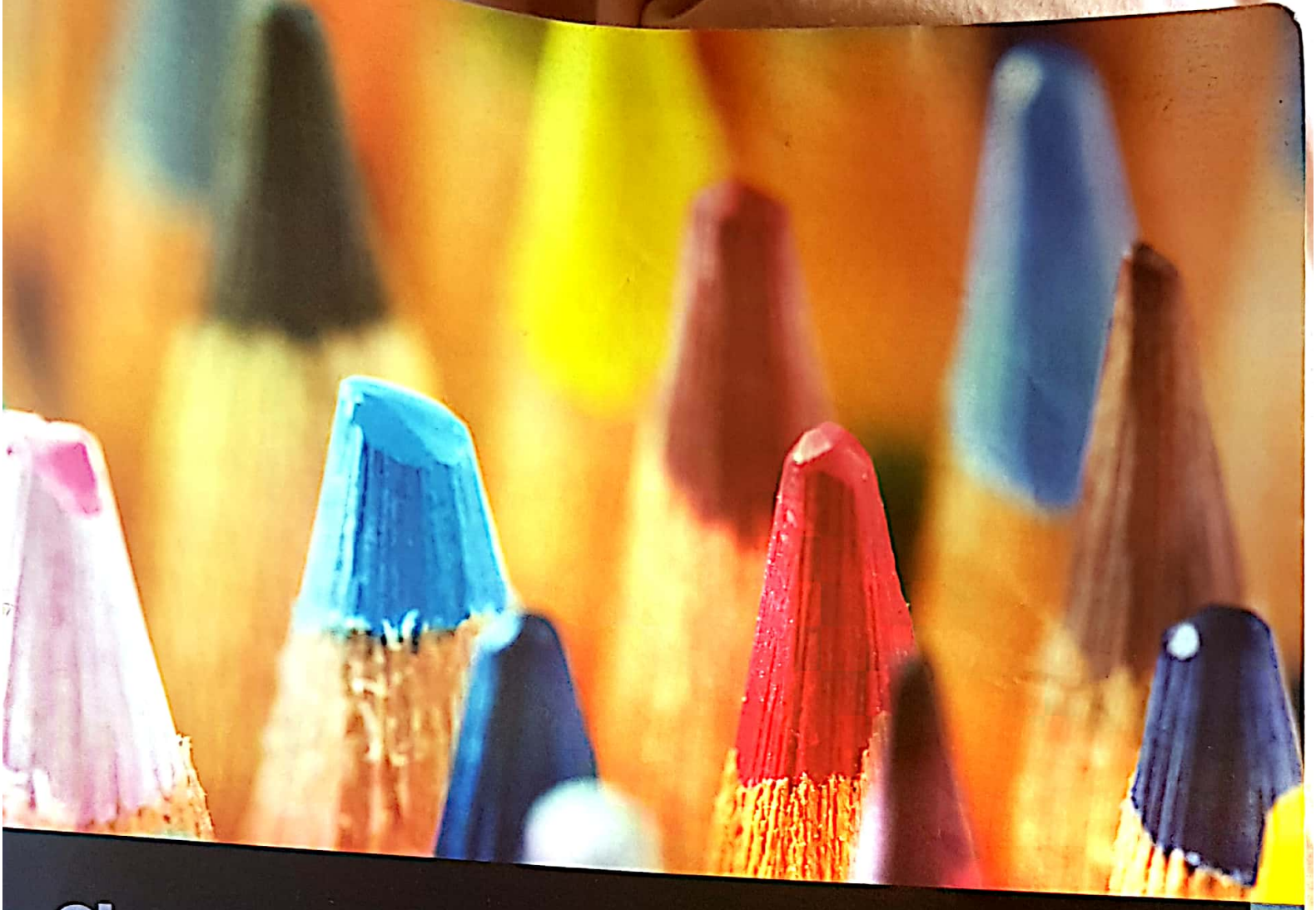
Coursebook



Contents

Issues and debates at AS Level

Issues and debates at A Level



Chapter 1
Research methods

Introduction

Psychology is a science, so the way psychological phenomena are explored is a research process. The methods used to investigate questions in psychology are called 'research methods'. This chapter will help you to understand how those methods are used by psychologists to find out about human (and animal) cognition, emotions and behaviour.

The chapter is divided into several sections, covering the basic research methods that you need to understand: experiments, self-reports, case studies, observations and correlations.

In addition, you will learn about features of the research process (hypotheses, variables, designs and sampling) and data and data analysis. There are also two further topics, which you will also consider within issues and debates: ethical and methodological issues. Together, these will help you to understand and be able to evaluate all aspects of research methods and to be able to apply your knowledge of research methods to novel research situations.

Why do psychologists do research?

As students, you may be bombarded with 'facts' about how to improve your learning. Perhaps you have heard of different learning styles, or the benefits of repetition or mind maps to help you to revise. Each of these methods should have been tested to see if they actually work (although many haven't!). The process of research allows scientists such as psychologists to test ideas in order to discover whether there is evidence to support them. This is how we decide which drugs or therapies work best for mental illnesses, whether different displays or music help to sell products in shops, and how we should organise work schedules to help factory workers to be efficient and healthy.

To be trustworthy, research needs to be planned well and conducted effectively. Imagine an investigation into new classroom techniques. If the researcher didn't know how hard the children worked, and compared the new techniques on a lazy class and a highly motivated class, this would produce false results. Consider a study into consumer psychology that compared how many goods were sold with and without music playing in the store. The researcher only played music at the weekends and played no music on weekdays. Would you believe the findings of studies such as these?

1.1 Where do we focus when we concentrate on a problem?

Reflections: Next time you see someone thinking really hard, perhaps trying to remember a name or work out the answer to a question, watch their eyes. It has been suggested that in such situations our eyes tend to look upwards and to the left (Figure 1.1). Consider how you might test whether this is true. Would you wait for people to get confused and then look at what they do, or would you give them a puzzle to make them think? How would you decide where they are looking? What would you do to be sure that they aren't just looking around the room for clues? Being able to decide on the answers to questions such as these is the basis of designing experiments in psychology.

1.1 Experiments

An **experiment** is an investigation which is looking for a cause-and-effect relationship. The researcher investigates the way one variable, called the **independent variable**, is responsible for the effect in another, the **dependent variable**. To test this, the researcher manipulates the independent variable (IV) to produce two or more conditions, such as 'high' or 'low' light levels or 'early' and 'late' in the day. The effect of these conditions on the dependent variable (DV) is measured. For example, an IV of light level might affect attention, with people being better at paying attention when the light levels are high. How well people pay attention would be the DV. If there is a big difference in the DV between the conditions, the researcher would conclude that the IV has caused the difference in the DV, i.e. that light levels affect attention (Figure 1.2).

KEY TERMS

experiment: an investigation looking for a causal relationship in which an independent variable is manipulated and is expected to be responsible for changes in the dependent variable.

independent variable: the factor under investigation in an experiment which is manipulated to create two or more conditions (levels) and is expected to be responsible for changes in the dependent variable.

dependent variable: the factor in an experiment which is measured and is expected to change under the influence of the independent variable.

1.2 An experiment can investigate whether the light level affects how well we concentrate

In order to be more certain that the difference between the conditions is caused by the IV, the researcher needs to control any other variables that might affect the DV. For example, people might find it harder to be attentive if they have eaten, exercised or sat through a very dull class. Such **extraneous variables** should therefore be controlled, i.e. kept the same in each condition (or 'level of the IV').

The levels of the IV being compared may be two or more **experimental conditions** (such as bright and dull artificial lights) or there may be one or more experimental conditions which are compared to a **control condition** (for example, artificial light compared to daylight). The control condition is simply the absence of the experimental variable. For example, in a comparison of the effect of eating chocolate on paying attention, we might compare either the effect of eating one bar or two bars (two experimental conditions) or the effect of eating one bar to no chocolate at all (one experimental and one control condition).

KEY TERMS

extraneous variable: a variable which either acts randomly, affecting the DV in all levels of the IV or systematically, i.e. on one level of the IV (called a confounding variable) so can obscure the effect of the IV, making the results difficult to interpret.

experimental condition: one or more of the situations in an experiment which represent different levels of the IV and are compared (or compared to a control condition).

control condition: a level of the IV in an experiment from which the IV is absent. It is compared to one or more experimental conditions.

laboratory experiment: a research method in which there is an IV, a DV and strict controls. It looks for a causal relationship and is conducted in a setting that is not in the usual environment for the participants with regard to the behaviour they are performing.

RESEARCH METHODS IN PRACTICE

A researcher might conduct a **laboratory experiment** to test the effect of the **independent variable** of time of day on the **dependent variable** of happiness of students. They might choose to control **extraneous variables** such as which lessons the students were in and whether they had recently eaten since these might affect happiness too. This would be a comparison between two **experimental conditions**.

Reflections: Look at the Research methods in practice box. Can you suggest:

- two different times of day to use as the levels of the *independent variable*
- how the *dependent variable* might be measured
- one other *extraneous variable* that it would be important to control?

Experimental design

The way that participants are used in different levels of the IV is called the **experimental design**. They may be allocated to all, or only one, of the levels of the IV.

The three experimental designs are:

- independent measures design
- repeated measures design
- matched pairs design.

Independent measures design

In an **independent measures design**, a separate group of participants is used for each experimental condition or level of the IV. This means that the data for each level of the IV is 'independent' because it is not related to any other data – it has come from different people. Note that this is a different use of the word 'independent' from that in the 'independent variable'.

If we wanted to know whether seeing aggressive models on television had long-term effects, we could (rather unethically) expose a group of young people to aggressive television and then wait for them to grow older. However, it would much quicker to compare two groups of adults, one group who had been allowed to watch aggressive TV as children and one group who had not been allowed to. This second example would be an independent measures design.

This design is good because the participants only encounter the experimental setting once. They are therefore unlikely

KEY TERMS

experimental design: the way in which participants are allocated to levels of the IV.

independent measures design: an experimental design in which a different group of participants is used for each level of the IV (condition).

to notice or respond to clues that might tell them the aims of the experiment (**demand characteristics**). One problem is that there might be individual differences between participants that could influence the findings. For example, in a study on the effect of noise on dreams, all the people who normally remember their dreams well might end up in the 'no noise' group. If so, it might look as though noise prevented dream recall when in fact it had little effect. This risk can be reduced by the **random allocation** of participants to different conditions. This spreads possible differences between individuals across the levels of the IV. To randomly allocate participants, each person is given a number, and the numbers are then randomly divided into two groups. This can be done by telling each participant a number, putting numbers into a hat and drawing out two sets, or using a random number generator (e.g. on a computer) to do the same thing.

Repeated measures design

In a **repeated measures design** the same group of people participate in every level of the IV. To help you to remember, think of the participants 'repeating' their performance under different conditions. For example, in a study looking at the effects of doodling on learning, we could count the number of words recalled in the same group of people when they did doodle and when they did not.

The main advantage of a repeated measures design is that each person acts as their own baseline. Any differences between participants that could influence their performance and therefore the DV will affect both levels of the IV in the same way. Individual differences are therefore unlikely to bias the findings. Imagine that in our experiment on doodling, one person was generally very quick to learn and another quite slow. In an independent measures design this might cause a problem if they were in different groups, but using a repeated measures design makes the differences between them less important, as both could show an improvement with doodling. Individual differences between participants are called **participant variables**. These variables, such as age, gender, personality or intelligence, can affect scores on the DV. It is therefore important to make sure that these variables do not hide, or exaggerate, differences between levels of the IV.

As each individual participates in every level of the IV they will perform the same or similar tasks two or more times. This can lead to a problem called the **order effect**. Repeated performance could cause participants to improve because they have encountered the task before – a **practice effect**. This matters because participants who were tested on a condition second would perform better than those who did it first. Alternatively, repetition might make performance worse, perhaps if they were bored or tired – a **fatigue effect**. In addition, the participants see both levels of the IV and have more opportunity to work out what is being tested, so are more likely to respond to demand characteristics.

Order effects can be solved in two ways: by randomisation or counterbalancing. Imagine an experiment with two conditions: learning while listening to music (M) and learning with no music (N). In randomisation, participants are randomly allocated to do either condition M followed by N, or vice versa. As some will do each order, any advantage of doing one of the conditions first will probably be evened out in the results. To be more certain that possible effects are evened out, **counterbalancing** can be used. Here, the group of participants is divided into two and one half will

KEY TERMS

demand characteristics: features of the experimental situation which give away the aims. They can cause participants to try to change their behaviour, e.g. to match their beliefs about what is supposed to happen, which reduces the validity of the study.

random allocation: a way to reduce the effect of confounding variables such as individual differences. Participants are put in each level of the IV such that each person has an equal chance of being in any condition.

repeated measures design: an experimental design in which each participant performs in every level of the IV.

participant variables: individual differences between participants (such as age, personality and intelligence) that could affect their behaviour in a study. They could hide or exaggerate differences between levels of the IV.

order effects: practice and fatigue effects are the consequences of participating in a study more than once, e.g. in a repeated measures design. They cause changes in performance between conditions that are not due to the IV, so can obscure the effect on the DV.

practice effect: a situation where participants' performance improves because they experience the experimental task more than once, e.g. due to familiarity or learning the task.

fatigue effect: a situation where participants' performance declines because they have experienced an experimental task more than once, e.g. due to boredom or tiredness.

randomisation:

counterbalancing: counterbalancing is used to overcome order effects in a repeated measures design. Each possible order of levels of the IV is performed by a different sub-group of participants. This can be described as an ABBA design, as half the participants do condition A then B, and half do B then A.

do M followed by N, the other half N followed by M. If on the second test there was a risk of participants accidentally including items learned in the first test, this would be a problem for exactly half the participants in the 'music' condition, and exactly half in the 'no music' condition. Alternatively, a different design could be used.

The problems associated with both independent measures and repeated measures designs are overcome in a **matched pairs design**. Participants are matched into pairs who are similar in ways that are important to the experiment, such as age, gender, intelligence or personality (Figure 1.3). This matching is done on variables relevant to the study, so in a study on the effects of playing a violent computer game, participants might be matched on their existing level of aggression. Identical twins make ideal matched pairs as they are both genetically the same and are likely to have had very similar experiences. Different groups of participants are then used for each level of the IV, with one participant from each pair being in each level of the IV. By using different participants in each group order effects are avoided and the matching of participants minimises the influence of individual differences.

KEY TERM

matched pairs design: an experimental design in which participants are arranged into pairs. Each pair is similar in ways that are important to the study and one member of each pair performs in a different level of the IV.

1.3 Identical twins are perfect participants for a matched pairs design

	Experimental design		
	Independent measures	**Repeated measures**	**Matched pairs**
Strengths	Different participants are used in each level of the IV so there are no order effects Participants see only one level of the IV, reducing the effect of demand characteristics Random allocation to levels of the IV can reduce the effects of individual differences	Participant variables are unlikely to distort the effect of the IV, as each participant does all levels Counterbalancing reduces order effects Uses fewer participants than repeated measures so is good when participants are hard to find or if participants are at risk	Participants see only one level of the IV, reducing the effect of demand characteristics Participant variables are less likely to distort the effect of the IV than in an independent measures design as individual differences are matched No order effects
Weaknesses	Participant variables can distort results if there are important individual differences between participants in different levels of the IV More participants are needed than in a repeated measures design so the study may be less ethical if participants are harmed and less effective if there is a small sample because participants are hard to find	Order effect could distort the results As participants see the experimental task more than once, they have greater exposure to demand characteristics	The similarity between pairs is limited by the matching process, so the right matching criteria must be chosen in advance for this to be effective Availability of matching pairs may be limited, making the sample size small (although some studies conducted on twins use very large numbers of pairs)

Table 1.1 Strengths and weaknesses of experimental designs

RESEARCH METHODS IN PRACTICE

A child psychologist conducted an experiment to look at the effect of violent computer games (Figure 1.4). There were two experimental conditions (violent and non-violent). The **dependent variable** was the children's subsequent violent behaviour. The **experimental design** chosen was an **independent measures design**, with different children in each of the experimental conditions. If a **repeated measures design** had been used, in which the same children played each type of game, there could be **order effects**. For example, aggression caused by playing the violent game could still affect children in the non-violent game condition if they did this second. If this were the case, the problem could be reduced by using **counterbalancing**.

1.4 Are children more violent after they have played a violent computer game than before?

However, the use of an independent measures design risks **participant variables**, such as the original level of violence of each child, affecting the results. This could be reduced by either using **random allocation** of participants to each condition or by using a **matched pairs design**. In this case, children with similar aggression levels would be put in the different conditions. To avoid **demand characteristics**, the children would ideally be unaware that they are in an experiment, perhaps by telling them that they are in a computer games competition.

KEY TERMS

standardisation: keeping the procedure for each participant in an experiment (or interview) exactly the same to ensure that any differences between participants or conditions are due to the variables under investigation rather than differences in the way they were treated.

reliability: the extent to which a procedure, task or measure is consistent, for example that it would produce the same results with the same people on each occasion.

Reflections: Look at the Research methods in practice box. Think about the following:

- It would be a good idea to have another level of the IV that did not use a computer game but did use a computer, such as looking at non-violent pictures. Would this be a control condition or another experimental condition?
- One potential order effect that could arise if a repeated measures design was used for this experiment is that the children might get fed up with playing computer games by the second condition. Is this a practice effect or a fatigue effect?
- Suggest a *participant variable* other than initial level of violence that could affect the results of this study.

Types of experiments

Laboratory experiments

Many experiments in psychology are conducted in artificial surroundings, such as a laboratory. Experiments conducted in this way are called laboratory experiments; the participants are not in their usual environment for the behaviour they are performing, and there are strict controls over the situation. For example, a laboratory experiment on the attention of schoolchildren in high and low light levels could be conducted. It might be investigated by testing the children on a computerised attention task conducted in a psychology room in a university.

Evaluating laboratory experiments

Laboratory experiments use many controls. In addition, researchers in laboratory experiments can use **standardisation**, which means that the procedure for each participant can be kept exactly the same. Both controls and standardisation help to make the findings of the experiment **reliable**, that is the researchers would be more certain that the procedures and measures they are using are consistent. Controlling variables also improves **validity** – how certain the researcher can be that they are testing what they claim to be testing. By keeping the situation the same, the researcher can be more certain that any differences in the DV really are due to the differences between levels of the IV rather than due to any extraneous variables.

KEY TERM

validity: the extent to which the researcher is testing what they claim to be testing.

RESEARCH METHODS IN PRACTICE

Dr Splash is conducting a laboratory experiment to test whether older adults detect emotions as quickly as younger people. He has two groups of participants, older and younger ones. This is his IV. He tests them by comparing how quickly they press a button to say that they have recognised the emotion on a face of a screen. This is the DV. Each participant sits at the same distance from the screen. This is one aspect of the **standardisation** of the procedure. In a pilot study, Dr Splash had shown the participants pictures and used a stopwatch to time their reactions himself, but he found he was not very consistent in his ability to stop timing exactly when the participant responded. He therefore changed to the computerised system to improve **reliability**. The pictures of faces included both younger and older people to ensure it was a **valid** test.

Reflections: Look at the Research methods in practice box above.

- Define the independent variable.
- Define the dependent variable.
- Name and explain the experimental design being used in this study.
- Suggest why the use of faces of a range of ages would have improved validity.

Field experiments

Returning to the idea at the beginning of this section of the effect of light levels, the schoolchildren could be tested by altering the number of lights turned on in their normal classroom. Light level would still be the IV and the levels of the IV could be 'all the lights on' and 'half the lights on'. The DV of attention could then be measured by looking at their scores on a class test they were due to take that day. This is still an experiment because it has an IV and a DV (and there will still be some controls, such as the amount of time they spend studying for the test). However, it would be a **field experiment** because the children are being tested on a usual behaviour (the topic test) in their normal environment (the classroom).

Evaluating field experiments

It is a little harder to control variables and standardise procedures in a field experiment than a laboratory experiment. Reliability and validity may therefore be lower. However, validity might be improved because the participants are performing a task that seems normal in a familiar environment. School students taken into a university laboratory might concentrate really hard because they are nervous or interested, which might cover up any differences between the different light level conditions. This means the findings from the laboratory would not **generalise** to other settings as well as those from the classroom. This is a problem of **ecological validity**, and field experiments often have better ecological validity than laboratory experiments (but not always).

Another advantage, if the participants are unaware that they are in an experiment, is that there may be fewer demand characteristics than there would be in a laboratory experiment. These are any features of the experiment that give away the aims and cause participants' behaviour to change, for example to try to 'make the experiment work'.

Natural experiments

A third type of experiment is the **natural experiment**. This is not a true experiment because the researcher cannot manipulate the levels of the IV. The differences or changes in the IV exist, or would occur, even in the absence of the experiment. For example, children's attention could be measured on very dull and very bright days, when the amount of light in the classroom differed (even with the lights turned on). The DV could again be measured with a class test.

KEY TERMS

field experiment: an investigation looking for a causal relationship in which an independent variable is manipulated and is expected to be responsible for changes in the dependent variable. It is conducted in the normal environment for the participants for the behaviour being investigated.

generalise: apply the findings of a study more widely, e.g. to other settings and populations.

ecological validity: the extent to which the findings of research in one situation would generalise to other situations. This is influenced by whether the situation (e.g. a laboratory) represents the real world effectively and whether the task is relevant to real life (has **mundane realism**).

natural experiment: an investigation looking for a causal relationship in which the independent variable cannot be directly manipulated by the experimenter. Instead they study the effect of an existing difference or change. Since the researcher cannot manipulate the levels of the IV it is not a true experiment.

Evaluating natural experiments

Using this method there is less opportunity to control and standardise the situation. There may be **uncontrolled variables**, such as how warm the classroom is. It might be much warmer on sunny days for example. This could matter because the warmth might make the children sleepy and less able to concentrate. This would lower the validity of the findings, although this is countered by the familiarity of the task and setting, which would increase ecological validity.

KEY TERM

uncontrolled variable: a confounding variable that may not have been identified and eliminated in an experiment, which can confuse the results. It may be a feature of the participants or the situation.

Types of experiment			
	Laboratory experiment	**Field experiment**	**Natural experiment**
Strengths	Good control of extraneous variables, raising validity Causal relationships can be determined Standardised procedures raise reliability and allow replication	As participants are in their normal situation for the activity being studied they are likely to behave naturally, making the results representative If participants are unaware that they are in a study, the problem of demand characteristics is less than in laboratory experiments	They can be used to study real-world issues If participants are in their normal situation, their behaviour is likely to be representative If participants are unaware that they are in a study, demand characteristics will be less problematic They enable researchers to investigate variables that it would not be practical or ethical to manipulate
Weaknesses	The artificial situation could make participants' behaviour unrepresentative Participants could respond to demand characteristics and alter their behaviour	Control of extraneous variables is harder than in laboratory experiments, lowering reliability and making replication difficult The researcher will be less sure that changes in the DV have been caused by changes in the IV than in a laboratory experiment Participants may be unaware that they are in a study, raising ethical issues	They are possible only when differences arise naturally Control over extraneous variables is often very difficult As the researcher is not manipulating the IV, they will be less sure of the cause of changes in the DV, so a causal relationship cannot necessarily be established They are often hard to replicate, as controls and standardisation are hard to implement, so the reliability may be low

Table 1.2 Strengths and weaknesses of experimental methods

RESEARCH METHODS IN PRACTICE

A research team is deciding how to test the effect of watching television on children's pro-social behaviour, that is, how nice children are to each other. They will measure pro-social behaviour by observing how often the children hold hands. They are considering two methods. One is a **field experiment**, in which parents either do or do not allow their child to watch television. Alternatively, they could observe the children in a remote place that has no television and then observe them again after the area has begun to receive satellite transmissions. This would be a **natural experiment**. Both of these studies would have more **ecological validity** than a laboratory experiment in which children were shown additional television, because in a laboratory the children would in an unfamiliar environment so may not pay attention to the television if they were nervous or distracted. In both situations there may be **uncontrolled variables**, such as which exact programmes were watched, and for how long. These factors could affect later pro-social behaviour. If the children are aware that their television viewing is being manipulated (in the field experiment) or their pro-social behaviour is being observed (in either experiment) they may try to alter their behaviour to meet the research team's expectations, for example being extra nice to each other (or especially nasty!).

Reflections: Look at the Research methods in practice box above. Which of the following can you identify?

- Independent variable
- Dependent variable.

Is there a *control condition*?

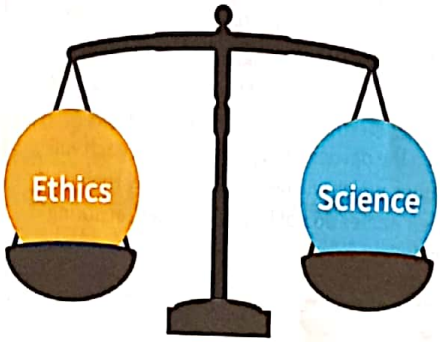
Can you suggest one *extraneous variable* that it would be important to control?

What effect might *demand characteristics* have in this study?

Suggest **one** strength and **one** weakness of conducting the study as a natural experiment in terms of *generalisability*.

Ethics in experiments

The role of ethics in psychology is discussed in detail in Section 1.10. Here we will briefly consider ethics in experiments (Figure 1.5). A participant in a laboratory experiment is likely to know that they are participating in a study and can readily be asked for their **informed consent**. However, it may be necessary to deceive them in order to avoid them working out the aim of the study and altering their behaviour, i.e. to reduce demand characteristics. There is therefore a balance between good ethics and good science. In field and natural experiments, in contrast, it may not be possible to gain consent as the participants may be unaware that they are even in a study. This is an ethical problem because participants should have the right to know what they are entering into and to agree to participate or not. They should also have the **right to withdraw**, which they cannot do if they do not even know that they are in a study, and they should be protected from possible harm.



1.5 Researchers must achieve a balance between good ethics and good science

In all experiments, privacy and confidentiality are important. **Privacy** can be respected in laboratory experiments because the tests or questions used are pre-planned. In the natural settings of field and natural experiments, however, there is a risk of invading privacy so researchers must be more careful of this. **Confidentiality** can be respected in all experiments by keeping the participants' data secure and anonymous, although if the participants are unaware that data has been collected, as in a field experiment, it is important to ensure that they cannot be individually identified, for example by their place of work.

KEY TERMS

informed consent: knowing enough about a study to decide whether you want to agree to participate.

right to withdraw: a participant should know that they can remove themselves, and their data, from the study at any time.

privacy: participants' emotions and physical space should not be invaded, for example they should not be observed in situations or places where they would not expect to be seen.

confidentiality: participants' results and personal information should be kept safely and not released to anyone outside the study.

RESEARCH METHODS IN PRACTICE

A psychology department ethical committee is looking at a research proposal for a study about the effect of cognitions on a therapy designed to help people to relax. The researchers only plan to ask for **consent** about the procedure they will use – listening to an imagery-based relaxation tape – and not their aim. They intend to deceive the participants about the independent variable, which will be either to tell them what will really happen – their pulse rate should fall – or to give them false information by telling them that some people see disturbing flashing lights. When the participants are given the limited information at the start of the study, they will also be told that they can leave at any time, giving them the **right to withdraw**. The instructions on the tape tell the participants to imagine relaxing, intimate thoughts. However, they will also be told that they will not be asked about these thoughts, which ensures their **privacy** is protected. When the participants join the study, each will be given a number, which will be used to identify their data so that their names do not have to be used, ensuring their **confidentiality**.

10 **Reflections:** Look at the Research methods in practice box above. Which of the following can you identify?

- The type of *experiment* being planned
- The independent variable
- The dependent variable
- The experimental design

Can you suggest one way in which possible harm to participants could arise as a result of this study?

Suggest why participants may want to withdraw from the study.

Why might it be necessary for the researchers to *deceive* the participants?

Applying your knowledge of experiments to novel research situations

You should be able to recognise experiments (including the IV and DV – and be able to operationalise them, i.e. define them in detail) to decide whether an experiment is a laboratory, field or natural experiment and to comment on controls, standardisation, ethics and reliability and validity. In addition, you should be able to plan an experiment, deciding on an IV and a DV, the type of experiment and how to implement suitable controls and to avoid ethical issues.

SELF-ASSESSMENT QUESTIONS

1 Barry and Anouk are deciding how to test whether gender affects artistic ability. Barry suggests doing a study in the psychology department where they ask students to come in for a study about memory in which they must redraw a complicated image. Barry and Anouk can then see how well they do it. Anouk thinks it would be better to persuade the art teacher to use an art class and set a lesson where students have to copy the same complicated image.

a Explain the type of experiment that is being suggested:

i by Barry

ii by Anouk.

b The independent variable is the same in Barry and Anouk's studies, as is the dependent variable.

i Describe the independent variable (IV).

ii Describe the dependent variable (DV).

c Explain **one** ethical issue that is clear from the procedure they have suggested.

d Suggest **one other** ethical issue and how they could avoid problems with this issue.

1.2 Self-reports

In a **self-report**, the participant gives the researcher information about themselves directly. This is different from experimental tests or observations where the researcher finds the data out from the participant. There are two techniques, questionnaires and interviews, both of which ask the participant questions.

Questionnaires

In a **questionnaire**, the questions are presented to the participant in written form. This may be on paper or as an online survey. There are several different types of questions. The two most important are **closed questions**, which have

KEY TERMS

self-report: a research method, such as a questionnaire or interview, which obtains data by asking participants to provide information about themselves.

questionnaire: a research method that uses written questions.

closed questions: questionnaire, interview or test items that produce quantitative data. They have only a few, stated alternative responses and no opportunity to expand on answers.

a fixed set of possible responses, and **open questions**, which ask for descriptive answers in the participant's own words. Closed questions can take the form of simple choices, such as those asking for yes/no answers or items from a list. Other forms of closed questions include rating scales (where a number is chosen, e.g. between 0 and 5) and Likert scales, which ask the respondent to say how much they agree with a statement such as 'Obesity is not important' or 'Exercise is a necessity' using the choices 'strongly agree / agree / don't know / disagree / strongly disagree'. Some examples of closed questions are as follows:

- What is your gender: male or female?
- How do you travel to school? walk / bicycle / bus / train / car
- Indicate which animal(s) scare you: dog, spider, cat, rat, fish, rabbit, bird. [You may tick as many as you like]
- How much do you like psychology on a scale of 0–4? (0 = not at all, 4 = very much)

Open questions prompt the respondent to give detailed answers, which may be quite long. They contain more depth than the answers to closed questions and are more likely to be able to explore the reasons behind behaviours, emotions or reasoning. They typically ask 'Why…' or simply 'Describe…'. Some examples of open questions are as follows:

- What do you think about children having access to the internet?
- Why do you believe it is important to help people who suffer from phobias?
- How would you suggest parents should discipline their children?
- When do you feel it is important to allow young people the freedom to control their own TV viewing?
- Describe your views on the use of social media sites with regard to encouraging helping behaviour.
- Explain how you would respond if you were told to hurt another person.

KEY TERMS

open questions: questionnaire, interview or test items that produce qualitative data. Participants give full and detailed answers in their own words, i.e. no categories or choices are given.

inter-rater reliability: the extent to which two researchers interpreting qualitative responses in a questionnaire (or interview) will produce the same records from the same raw data.

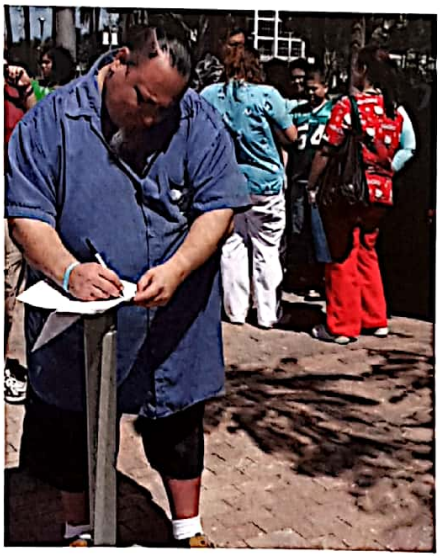
social desirability bias: trying to present oneself in the best light by determining what a test is asking.

filler questions: items put into a questionnaire, interview or test to disguise the aim of the study by hiding the important questions among irrelevant ones so that participants are less likely to alter their behaviour by working out the aims.

Evaluating questionnaires

Questionnaires using mainly closed questions are easier to analyse than interviews (using more open questions) as they can be used to produce totals of each category of answers so making it simple to summarise the findings. It is also possible to work out averages, which can help to describe the patterns in the results. Where qualitative data is gathered from questionnaires, it produces more detailed, in-depth information. This is an advantage, although it also leads to a problem. Answers to open questions have to be interpreted, and this can lead to a lack of reliability as the researcher may not be consistent in their interpretation. If more than one researcher is involved, there may also be differences between them. This would be a lack of **inter-rater reliability**.

One problem with questionnaires is that it is easy for participants to ignore them, which means the return rate may be very low. Importantly, the people who do reply to a questionnaire may all be quite similar, for example have time to spend because they are unemployed or retired. This would mean all the people who filled out the questionnaire would be quite similar.



1.6 People may lie in questionnaires, lowering validity, for example giving socially desirable responses to questionnaires about eating habits

Another problem with questionnaires is that participants may lie. They may do this because they want to look more acceptable; this is called a **social desirability bias** (Figure 1.6). Participants may also lie if they believe they have worked out the aim of the study. To avoid this, researchers sometimes include **filler questions** among the real questions. The answers to filler questions are not analysed in the research since they serve only to hide the real purpose of the study.

RESEARCH METHODS IN PRACTICE

Dr Blot is a psychology teacher. She wanted to know how her students were progressing on the course. She decided to use the **self-report** method and used an online **questionnaire** that the students did in their free time to collect data. This included several **closed questions** (1–4), which collected quantitative data, and some **open questions** (5 and 6) which collected qualitative data. She asked her colleague to help her to interpret the responses to the open questions and to help to ensure that they had good **inter-rater reliability**, she devised a list to help them to interpret questions 5 and 6. For question 5 it included looking for comments about:

reading up notes
copying up notes
reading the textbook
looking things up online
asking friends
checking with the teacher

For question 6 it included looking for comments about:

copying out notes
making summary notes
making mind maps
using past paper questions
making test cards

Some of the questions on the questionnaire were:

1 How often do you do the homework set?

always | *sometimes* | *never*

2 Have you written yourself a research methods glossary?

yes | *no*

3 'Psychology is a difficult subject'. Do you:

strongly agree | *agree* | *don't know* | *disagree* | *strongly disagree*

4 Rate from 0 to 6 how well you understand the topic we have just completed:

0 = don't understand at all

6 = completely understand

5 Explain what you do after each lesson to help you to remember what you have learned.

6 Describe how you will plan your revision for the next test.

Reflections: Look at the Research methods in practice box.

- Explain the difference between the *open* and *closed questions*.
- Suggest one more open question.
- Suggest one more closed question.
- Suggest why Dr Blot may have chosen to use an online questionnaire rather than one the students did on paper in the classroom.
- Explain why it was important that Dr Blot took steps to raise *inter-rater reliability*.

Interviews

In an **interview**, the researcher is typically face-to-face with the participant. Interviews can, however, be conducted through any medium that allows real-time interaction, such as by telephone or through a chat facility. The same kinds of questions can be asked in interviews as in questionnaires, although more open questions may be used.

The schedule of questions, that is the range of questions that are asked and the order of them, differs between different types of interviews. In a **structured interview**, the questions asked are the same for every participant and the order is fixed. There may even be instructions for the interviewer about how to sit or dress in order that the procedure is standardised each time data is collected. In an **unstructured interview**, in contrast, the questions asked depend on what the participant says, so the questions may be different for each participant. This is a very flexible technique but it may be hard to compare data collected from different participants or by different researchers. A compromise is a **semi-structured interview**. Here, there

KEY TERMS

interview: a research method using verbal questions asked directly, e.g. face-to-face or on the telephone.

structured interview: an interview with questions in a fixed order which may be scripted. Consistency might also be required for the interviewer's posture, voice, etc. so they are standardised.

unstructured interview: an interview in which most questions (after the first one) depend on the respondent's answers. A list of topics may be given to the interviewer.

semi-structured interview: an interview with a fixed list of open and closed questions. The interviewer can add more questions if necessary.

are some fixed questions, which make sure that there is some similar information from every participant. This means that comparisons can be made between them, and averages can be calculated if this is appropriate. In addition, it is possible to ask some questions that are specific to individual participants. This allows the researcher to develop ideas and explore issues that are particular to that person.

Evaluating interviews

As with questionnaires, interviewees may lie either because they want to seem more acceptable (a social desirability bias) or because they think they know the aim of the study, and are either trying to help the researcher by giving the answers they need, or to disrupt the research by doing the opposite. Interviewing is often time consuming and this can be a problem if it restricts the types of participants who volunteer for the research because it would give a narrow representation of feelings, beliefs or experiences.

When interpreting participants' responses to questions in an interview, researchers must be careful not to be **subjective**, that is, to produce findings which are based on a personal perspective. Instead, they should aim for **objectivity**, i.e. taking a view that is not led by one's own feelings or beliefs. To achieve this, the interviewer may ask other researchers, who are experienced but unaware of the aims of their research, to interpret the findings.

KEY TERMS

subjectivity: a personal viewpoint, which may be biased by one's feelings, beliefs or experiences, so may differ between individual researchers. It is not independent of the situation.

objectivity: an unbiased external viewpoint that is not affected by an individual's feelings, beliefs or experiences, so should be consistent between different researchers.

Applying your knowledge of self-reports to novel research situations

You should be able to recognise self-report studies, and decide whether they are questionnaires or interviews. You should also be able to choose which of these to use in a new situation. In addition, you should be able to recognise and write different types of questions (open and closed) and to identify and design different interview schedules (structured, semi-structured and unstructured). When doing this, it is important to consider how the method used affects the availability of different types of participants and their honesty, as this affects the validity of the findings.

You should also think about the kinds of data that are produced, and the way it will be used. Although numerical data from closed questions can be analysed mathematically, data from open questions provides more in-depth information which may be more valid. For example, a closed question might not have a response close to a person's view, so an open question would allow that person to express views that they could not do in the choices available in the closed question.

Finally, the reliability of self-report data is important. Questionnaires and structured interviews may be higher in reliability because they are likely to be administered in a consistent way and because they generate numerical results which do not need interpretation. Responses to open questions, in contrast, have to be interpreted by the researcher and since they may differ in their opinions there is the possibility that they will be subjective.

RESEARCH METHODS IN PRACTICE

Dr Splash is planning an **interview**-based study because he wants to confirm that a new shopping centre is making people more helpful to each other. He wants to collect objective data about the number of times people are altruistic so has devised a **structured interview** with a list of specific questions such as 'How many times has someone held a door open for you?', 'Have you helped anyone carry their shopping?' and 'Have you seen anyone assisting a parent with a buggy?'. However, he is worried that this may produce very limited data so has an alternative plan to use an **unstructured interview**. This would begin with the question 'Please can you describe how friendly or helpful you have found people to be at the new shopping centre', after which he would base his questions on what they said. A colleague suggests that both methods have limitations. Interpreting the responses to the unstructured interview might lead to very **subjective** data, especially as Dr Splash already believes that the participants will be finding the shopping centre encourages helpfulness. Although the data from the closed questions in the structured interview might produce more objective measures, this would limit opportunities for asking participants to expand on their answers. The colleague suggests that a **semi-structured interview** might be better.

Reflections: Look at the previous Research methods in practice box.

- Why is the first of Dr Splash's suggestions a structured interview?
- Why would the data from these questions be more objective?
- Why is the second plan an unstructured interview?
- What is the problem with subjective interpretations of the participants' responses in the unstructured interview?
- Suggest why a semi-structured interview would be better in this case.

SELF-ASSESSMENT QUESTIONS

2 Shareen and Judith are investigating people's phobias. They have decided to use self-reports. Shareen is suggesting using a questionnaire and Judith wants to interview people instead.

a Suggest **one** closed question and **one** open question that Shareen could use.

b Suggest **one** reason why Judith might want to conduct an unstructured interview.

c Describe **one** ethical problem that might arise in **either** Shareen's **or** Judith's version of the study.

1.3 Case studies

A case study is a detailed investigation of a single instance, usually just one person, although it could, for example, be a single family or institution. The data collected is detailed and in-depth and may be obtained using a variety of different techniques. For example, the participant may be interviewed, observed, given tests or asked to fill in questionnaires. Case studies are particularly useful for looking at rare cases where a detailed description is useful, and for following developmental changes, where the progress of a child, or a person with a disorder can be tracked through their improvement or decline. Case studies are therefore sometimes linked to therapy but it is important to remember that when the case study as a research method is being discussed, the therapeutic purpose is not the main aim.

Evaluating case studies

In some ways, the findings from case studies are highly valid, as the individual is explored in great depth and within a genuine context such as their work or family. Validity may be improved further using triangulation, where the use of different techniques should produce similar findings, for example observations and interviews with the participant and questionnaires for their family should all lead to similar conclusions. The research includes details such as their past as well as their present situation, their social interactions, their thinking and their emotions as well as their behaviours. Such detail, however, carries risks. One potential problem is the development of a close relationship with the researcher. This may make the researcher subjective in their outlook, which would reduce the validity of the study. The level of detail can also be an ethical threat, as the questions asked may intrude into the participant's private life and they may feel unable to refuse to answer them. The detail about the individual may make it hard to disguise their identity, even if they are not referred to by name, which would risk breaking the guideline of confidentiality.

Reliability is also an issue, as there is a single participant and perhaps one or only a few researchers. This, and their involvement with the case, means that they may find it hard to be objective, that is to take an external, unbiased view of the findings, for example when they interpret what the participant has said. This means that the findings may be limited to only this case, or to very few others.

Applying your knowledge of case studies to novel research situations

You should be able to recognise case studies, and when it is appropriate to use one. You should also be able to suggest possible techniques that could be used in a case study. When making these decisions, it is important to consider the validity and reliability of the findings. One way that the validity can be improved is though triangulation, where different methods are used within the case study to obtain the same information: for example, finding out about the participant's behaviour by observing them, interviewing them and asking their relatives to fill out a questionnaire. If the same results are obtained by all the methods, this suggests that the results are valid. Another consideration in planning case studies is an ethical one. Participants should be aware of their commitment, so that they can give their informed consent, and particular attention should be paid to ensuring their privacy is not invaded and that confidentiality is maintained.

RESEARCH METHODS IN PRACTICE

A psychologist in a sleep clinic has been conducting a **case study** on a patient, SL, who has had very bad dreams for several years. The psychologist uses an EEG to follow SL's sleep cycles and to detect when the patient is dreaming. The patient is then woken up and asked what the dream is about. SL has also been asked to keep a dream diary to record when the bad dreams occur. Members of SL's family have been interviewed to find out when the problems with nightmares started and how often they occur.

Reflections: Look at the Research methods in practice box above.

- How many different methods can you identify?
- Give two reasons why this is a case study.
- The psychologist is concerned that if the patient SL wanted to *withdraw* from the study, this might be quite difficult. Explain why this might be so.
- Explain the *ethical* reason for the researcher referring to the patient as SL.
- Suggest one **other** ethical issue that might be a problem in this study.
- Suggest **one** practical issue that might be a problem in this study.

SELF-ASSESSMENT QUESTIONS

3 Damon and Inka are planning a case study to investigate responses to emotional situations.

a Suggest **three** techniques that Damon and Inka might use in their case study.

b i Describe **two** ethical problems that might arise in the study that Damon and Inka are planning.

ii For **one** of these problems, suggest a possible solution.

c Explain whether the results from Damon and Inka's study would be typical of the way everyone would respond to emotional situations.

1.4 Observations

Observations involve watching human or animal participants. This can be done in two overall ways. A **naturalistic observation** is conducted in the participants' normal environment, without interference from the researchers in either the social or physical environment. A **controlled observation** is conducted in a situation which has been manipulated by the researchers. This may be in terms of the social or physical environment. Controlled observations can be done in either the participants' normal environment or in an artificial situation such as a laboratory.

At the beginning of a study, observations may be non-focused, that is, the whole range of possible behaviours are considered. If this continues throughout the study, it is called an **unstructured observation**. Usually, however, the range of behaviours studied is narrowed to a set of behaviours, and this technique is called a **structured observation**. The specific activities to be recorded are clearly defined in **behavioural categories**. This helps the observers to be consistent, i.e. it improves **inter-observer reliability**.

Another decision to be made is the role of the observer in the social setting. This may be participant or non-participant. A **participant observer** is part of the social setting, whereas a **non-participant observer** does not become involved in the situation being studied. This can be

KEY TERMS

naturalistic observation: a study conducted by watching the participants' behaviour in their normal environment without interference from the researchers in either the social or physical environment.

controlled observation: a study conducted by watching the participants' behaviour in a situation in which the social or physical environment has been manipulated by the researchers. It can be conducted in either the participants' normal environment or in an artificial situation.

unstructured observation: a study in which the observer records the whole range of possible behaviours, which is usually confined to a pilot stage at the beginning of a study to refine the behavioural categories to be observed.

structured observation: a study in which the observer records only a limited range of behaviours.

behavioural categories: the activities recorded in an observation. They should be operationalised (clearly defined) and should break a continuous stream of activity into discrete recordable events. They must be observable actions rather than inferred states.

inter-observer reliability: the consistency between two researchers watching the same event, i.e. whether they will produce the same records.

participant observer: a researcher who watches from the perspective of being part of the social setting.

non-participant observer: a researcher who does not become involved in the situation being studied, e.g. by watching through one-way glass or by keeping apart from the social group of the participants.

KEY TERMS

overt observer: the role of the observer is obvious to the participants.

covert observer: the role of the observer is not obvious, e.g. because they are hidden or disguised.

achieved by watching through one-way glass or by keeping apart from the social group of the participants.

The role played by the observer may be either **overt** (it is obvious that they are an observer) or **covert** (they are hidden or disguised so the participants do not know the individual is an observer). Participant observers are overt, for example, if the researcher is holding a clipboard. When a participant observer is disguised as a member of the social group (Figure 1.7), or when a non-participant observer is physically hidden (e.g. by using CCTV), they are covert. Participants cannot be aware that they are being watched if the observer is covert. This increases validity as it is unlikely that participants would be affected by being observed, so demand characteristics and the effects of social desirability are reduced. However, covert observations raise practical issues, as the observer must be either hidden, far away or disguised in their role. This may make data collection more difficult, potentially reducing validity and reliability. Furthermore, covert participant observation raises ethical issues as the participants cannot give informed consent, and if they work out the observer's role this can cause distress.

1.7 A covert participant observer is disguised by being part of the social group: which one is the observer?

Evaluating observations

Naturalistic observations have the advantage that the behaviours seen are true to life. They are more likely to reflect the way the individuals really behave than if there is interference in the situation from researchers, as is the case in controlled observations. However, there is no guarantee that the behaviours being studied will actually occur in a naturalistic situation, so it may be necessary to use a controlled observation.

Using an unstructured observation ensures that any important behaviours are recognised, but it may be very difficult to record all the activities accurately and many may be irrelevant. It is therefore likely, especially when only specific activities are of interest, that a structured observation will produce more reliable data.

The role played by the observer, and the participants' awareness of this, affect validity. If participants are unaware of the observer, or can ignore them, their activities are more likely to reflect their normal behaviour. This means that covert observers would produce more valid results than overt ones. However, in the case of covert participant observers there is an ethical issue of deception. The participants may interact with the observer in ways that they might have chosen not to if they had been aware that the individual was a researcher. This could invade their privacy and cause distress. However, an overt observer is likely to alter the behaviour of the participants as they are aware that they are being watched. This would reduce the validity of the findings as the activities being recorded are less likely to reflect real-world behaviour.

Applying your knowledge of observations to novel research situations

Observations can be used either as a research method or as a means to collect data in other research methods such as case studies, experiments or correlations. You should be able to distinguish between these two situations. Observation is being used as a research method in itself when the study consists solely of a means to collect data by watching participants and recording their behaviour directly to provide data. Observations are used as a technique to collect data about variables in other research methods when they are used to measure the dependent variable in an experiment or one or both variables in a correlation. In a case study, observations can be used alongside other techniques to explore a single instance in detail.

You will need to be able to decide when it is appropriate to use observations as a method, or as a technique within other methods. You should be able to recognise and justify choices about naturalistic versus controlled, structured versus unstructured, participant versus non-participant and covert versus overt observations. You will also need to be able to suggest ways of achieving these, such as how to make an observer participant or covert.

RESEARCH METHODS IN PRACTICE

Dr Blot is interested in whether her students detect each other's emotions and plans a **controlled observation**. She asks three students to act as confederates. They are told to take it in turns to appear quite sad in the common room at lunchtime. Dr Blot is on lunch duty with a colleague so they can act as **non-participant, overt** observers as they walk through the common room. The students will take no notice of them as they are used to them being there. Dr Blot suggests that she and her colleague use a list of specific behaviours to record, so they are doing a **structured observation**. This will also help to raise **inter-observer reliability** as they will be working from the same definitions, such as recording 'shows concern without action' if people look at the confederate without moving towards them, 'verbalises concern' if someone goes up to the confederate and asks them if they are OK and 'takes action' if they engage in a behaviour such as putting their arm round the confederate or buying them a drink.

Reflections: Look at the Research methods in practice box above.

- Suggest **one** other *behavioural category* that might have been included in the structured observation and define it.
- If Dr Blot had conducted an *unstructured observation*, how would the method have differed?
- An alternative plan would have been to have conducted a *naturalistic observation* and watched to see if any students appeared to be sad and how others responded. Outline **one** way it would have been ethically more acceptable and **one** way in which it would have been ethically less acceptable.
- Suggest how Dr Blot could have used a *participant, covert observer* rather than being an overt observer.

SELF-ASSESSMENT QUESTIONS

4 Debra and Jin want to use observations to find out about the behaviour of animals. Debra wants to go to the park and hide in a tree to observe the animals that live there. Jin thinks it would be better to set up an artificial situation and watch laboratory rats interacting with objects they would put in a special box.
   - **a** Who is suggesting a naturalistic observation and who is suggesting a controlled observation?
   - **b** Explain whether the observers in Debra's study would be overt or covert.
   - **c** For either Debra's or Jin's suggestion, decide whether it should be conducted as an unstructured or a structured observation and justify your choice.

1.5 Correlations

A correlational analysis is a technique used to investigate a link between two measured variables. **Correlations** are useful when it is possible only to measure variables, rather than manipulate them, i.e. when an experiment cannot be conducted. This may be because changing the variables would not be practical or would be unethical. For example, it would not be practical to conduct an experiment which controlled children's long-term exposure to television and it would not be ethical to increase real-life exposure to violent television programmes. Both of these could, however, be investigated using correlations. It is important to recognise that any link found between two variables in a correlation cannot be assumed to be a **causal relationship**, that is, we cannot know whether the change in one variable is *responsible* for the change in the other variable (Figure 1.8).

1.8 A bizarre positive correlation has been reported between ice cream consumption and murder rates. This relationship is a correlation, however, so we cannot conclude that eating ice cream causes people to commit murder

To look for a correlation between two variables, each variable must exist over a range and it must be possible to measure them numerically. Several techniques can be used to collect data for correlations, such as self-reports, observations and different kinds of tests.

We cannot say from one correlation that an increase in one variable has caused an increase (or decrease) in the other, because it is possible that the changes in both variables could be the result of another factor. Imagine that two variables are being measured: attention in class and score

KEY TERM

correlation: a research method which looks for a causal relationship between two measured variables. A change in one variable is related to a change in the other (although these changes cannot be assumed to be causal).

on a test. If these two correlate it would be tempting to say that paying attention in class is responsible for good test results but we cannot be sure of this. It is possible that both of these factors depend on another variable, such as the dedication of the individual student. The sort of student who pays more attention in class might also study much harder for the test. All we can conclude is that the two factors we have measured vary together, not that there is a cause-and-effect or causal relationship between them. As a consequence, it is important that you refer to 'measured variables' or 'co-variables' in a correlation and not independent and dependent variables. To make judgements about causality, an experiment must be used, so that we can be more certain that it is the manipulation of one variable that is responsible for the change in the other. If, on the other hand, we conduct a correlational study and find that there is *no* link between two variables, then we can conclude that there is no causal relationship.

The nature of the relationship between the two variables in a correlation can be described in terms of its *direction*. In a **positive correlation**, the two variables increase together. The change is in the same direction, so higher scores on one variable correspond with higher scores on the other. For example, in a positive correlation between exposure to aggressive models and violent behaviour, greater exposure to models would be linked to higher levels of violence. When two variables are **negatively correlated**, higher scores on one variable correspond with low scores on the other. For example, a negative correlation might exist between number of years in education and level of obedience: people with fewer years of education are more obedient (see also Section 1.9 on how to draw a scatter graph and a discussion of the *strength* of a correlation).

Evaluating correlations

A correlational study can only be valid if the measures of both variables test real phenomena in effective ways. To achieve this, the variables must be clearly defined and relate directly to the relationship being investigated. The reliability of a correlation depends on the measures of both variables being consistent. So, for some correlations, such as those using scientific scales (such as volume in cm^3 or time in seconds), the measures will be highly reliable. In other cases, such as studies correlating variables measured using self-reports or observations, there is a risk that reliability will be lower. This is because results from these measures may be less objective than from scientific measurements.

The main issue with correlations, however, is to remember that the conclusions do *not* necessarily reflect a causal relationship.

Applying your knowledge of correlations to novel research situations

Correlations provide a good starting point for research. They can indicate whether a relationship exists that might be worth pursuing with other research methods, such as experiments. Correlations are also useful because they enable researchers to explore problems when it is not practically or ethically possible to conduct experiments. You should be able to distinguish between correlations (with two measured variables) and experiments (where there is one variable – the IV – that is manipulated by the researcher and only one that is measured – the DV). You should also be able to recognise the difference between positive and negative correlations.

You will need to be able to decide when it is appropriate to use a correlation rather than any other method, for example when it is impossible to manipulate variables for practical or ethical reasons. You should also be able to justify choices about ways to measure the variables in a correlation and to suggest whether you would expect a positive or a negative correlation in a study. Finally, you will need to understand how to display the results of a correlational study on a scatter graph. This is discussed in Section 1.9.

KEY TERMS

positive correlation: a relationship between two variables in which an increase in one accompanies an increase in the other, i.e. the two variables increase together.

negative correlation: a relationship between two variables in which an increase in one accompanies a decrease in the other, i.e. higher scores on one variable correspond with lower scores on the other.

RESEARCH METHODS IN PRACTICE

Professor Smudge is studying phobias. She thinks that there may be a **correlation** between how long a phobia has lasted and how severe it is. She is asking her sample of participants with phobias to record how many years they have suffered with their fear and rate how much the phobia interferes with their life on a scale of 1 (hardly at all) to 10 (almost constantly and prevents me from functioning normally). If there is a link between the two measured variables, there are two possible outcomes. There may be a **positive correlation** – phobias that have lasted longer may be more severe. There may, however, be no relationship between the two variables. If so, when she plotted the results on a **scatter graph**, the points would appear randomly placed, rather than lying on a line.

Reflections: Look at the Research methods in practice box above.

- An alternative outcome might have been that phobias that have lasted longer are less severe. Explain why this is a negative correlation.
- If Professor Smudge found a positive correlation, it would be tempting, but incorrect, to say that the passage of time makes phobias worse. Why would this conclusion be incorrect?

SELF-ASSESSMENT QUESTIONS

5 Ekua and Takis are going to find out if there is a correlation between the amount of coffee people drink and the number of dreams they recall.

a Explain why this is a correlational study and not an experiment.

b Suggest whether the results will show a positive correlation or a negative correlation.

c i Suggest **one** way to measure the amount of coffee that people drink.

ii Explain either **one** advantage or **one** disadvantage of the way you have suggested measuring this variable.

1.6 Research processes

We began this chapter with a discussion about the need for research to test different ways to help students to learn or different therapies to help people who are mentally ill. These are examples of real-world problems that psychologists try to solve through their research. In this section we will consider the steps a psychologist might take in developing research to investigate a question or problem. This research process can be thought of as having several steps, the:

- development of an aim and **hypothesis**
- selection of a research method and, in an experiment, the experimental design
- definition, manipulation, measurement and control of variables
- ethical considerations
- selection of participants
- analysis of data, including the drawing of conclusions
- evaluation of research.

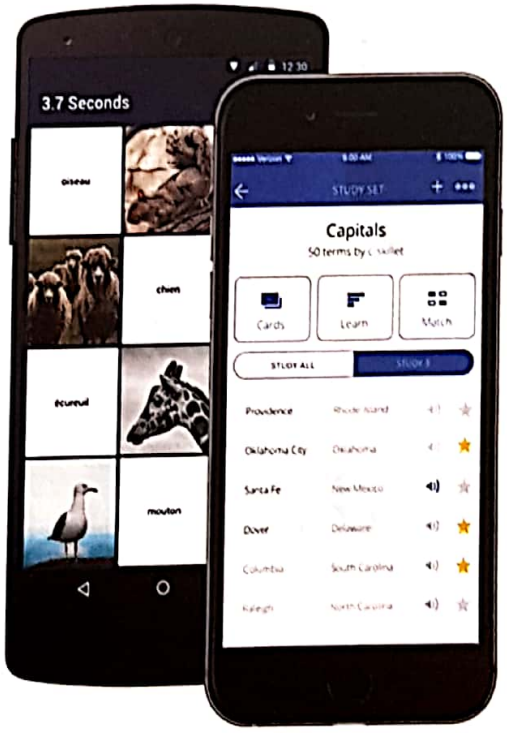
We will consider each of these steps in turn, evaluating the alternatives where appropriate and illustrating how the deas can be applied to novel research situations.

Aims and hypotheses

Aims

Consider the idea of different ways to help students to study, perhaps using mind maps or revision apps (see Figure 1.9). Imagine that a psychologist, Dr Blot, asks a few of her psychology students which method they prefer, and finds that both are quite popular. Dr Blot wants to know which is most effective. This is Dr Blot's aim – *to investigate whether mind maps or revision apps are more effective at helping students to learn.* So, the *aim* tells you the purpose of the investigation. It is generally expressed in terms of what the study intends to show.

In a correlation, the aim is to investigate a link or relationship between two measured variables, such as between the number of computer games a student plays and their final A Level grade.



1.9 Revision apps such as Quizlet and Gojimo may – or may not – help students to learn

Hypotheses

To make her research more exact, Dr Blot needs to present this aim as a hypothesis, that is, as a testable statement. A hypothesis should provide a little more detail about the

KEY TERM

hypothesis (plural hypotheses): a testable statement predicting a difference between levels of the independent variable (in an experiment) or a relationship between variables (in a correlation).

variables being investigated than the aim. Importantly, a hypothesis should also be *falsifiable*, that is it should be possible for it to be shown to be wrong. The main hypothesis in a study (sometimes called the **alternative hypothesis**) can be written in several different ways. They differ in terms of the nature of the prediction they make about the results of an investigation.

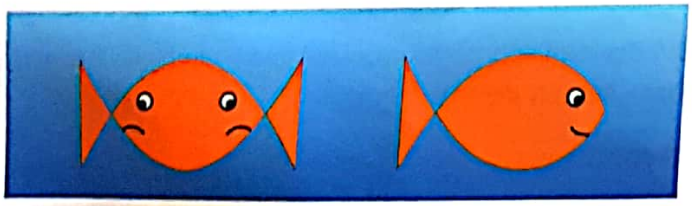
Non-directional hypotheses

A **non-directional (two-tailed) hypothesis** predicts that there will be an effect, but not the direction of that effect (Figure 1.10). In an experiment, this means that the hypothesis will suggests that the IV will change the DV but not whether the effect will be an increase or a decrease. This type of hypothesis is chosen if the effect of the variable is being tested for the first time, so there are no previous results to suggest what the results might be. For example, Dr Blot's hypothesis could be: *There is a difference between the effectiveness of mind maps and revision apps in helping students to learn.* Note that it is predicting a difference, but not which condition will be better at helping with learning.

A non-directional hypothesis in a correlational study predicts that there will be a relationship between the two measured variables. For example, a directional hypothesis might be: *There will be a correlation between the number of computer games a student plays and their final A Level grade.*

Directional hypotheses

When most previous research or other evidence suggests the nature or 'direction' of an effect we can use a **directional (one-tailed) hypothesis**. In an experiment this means saying which condition will be 'best' (i.e. produce the 'highest' scores) and in a correlational study, whether there will be a positive or negative correlation.



1.10 Unlike with a one-tailed hypothofish, you can't see which way a two-tailed hypothofish will swim

Returning to Dr Blot's study, there might be evidence that revision apps are better than mind maps, perhaps because they are more 'active' and being actively engaged helps memory. This is a directional prediction so the hypothesis might be: *Students using revision apps will learn better than students using mind maps.* Note that the opposite prediction could also be expressed as a directional hypothesis. This would be: *Students using mind maps will learn better than students using revision apps.* We might make this prediction if we believed that writing a mind map yourself was more effective than just reusing ready-made materials on revision apps.

A directional hypothesis for the correlational study about computer games and grades could say: *There will be a negative correlation between the number of computer games a student plays and their final A Level grade.* We might make this prediction if we believed that the time spent playing games might stop students working. However, a different directional hypothesis could be: *As the number of computer games a student has increases, their A Level grade increases.* We might make this prediction if we believed that students who engaged more with technology, even through games, were also more likely to benefit from technology-based learning aids. Remember that your hypothesis should not say that one factor *causes* the change in the other.

Null hypotheses

The alternative hypothesis is an alternative to the **null hypothesis**. In an experiment, the null hypothesis states that any difference in the DV between levels of the IV is so small that it is likely to have arisen by chance. For Dr Blot's study, the null hypothesis could be written either as: *There will be no difference between the effectiveness of mind maps and revision apps in helping students to learn* or *Any difference in effectiveness of mind maps and revision apps in helping students to learn is due to chance.*

KEY TERMS

alternative hypothesis: the testable statement which predicts a difference or relationship between variables in a particular investigation.

non-directional (two-tailed) hypothesis: a statement predicting only that one variable will be related to another, e.g. that there will be a difference in the DV between levels of the IV in an experiment or that there will be a relationship between the measured variables in a correlation.

directional (one-tailed) hypothesis: a statement predicting the direction of a relationship between variables, e.g. in an experiment whether the levels of the IV will produce an increase or a decrease in the DV or in a correlation whether an increase in one variable will be linked to an increase or a decrease in another variable.

null hypothesis: a testable statement saying that any difference or correlation in the results is due to chance, i.e. that no pattern in the results has arisen because of the variables being studied.

To help you to write null hypotheses for experiments, remember that they should say 'There will be no difference in the DV between *condition X* and *condition Y*' or that 'Any difference in the DV between *condition X* and *condition Y* is due to chance'. Make sure that you *always* state both of the levels of the IV and the DV otherwise your null hypothesis will not make sense. For example, the null hypothesis 'There is no difference between mind maps and revision apps' is meaningless.

Correlational studies also need a null hypothesis. These predict either no link or that any relationship could have occurred by chance. A general null hypothesis for a correlational study reads: *There will be no relationship between variable X and variable Y* (or *Any relationship between variable X and variable Y is due to chance*). For example: *There will be no relationship between the number of computer games a student has and their A Level grade* (or *Any relationship between the number of computer games a student has and their A Level grade is due to chance*).

Reflections: Look at the Research methods in practice box.

- Which study is an experiment and which is a correlation?
- Can you suggest a different directional hypothesis for the correlation, one that proposes a negative correlation?
- What is wrong with the alternative hypothesis '*Soft chairs will be better than hard chairs*'?

1.7 The definition, manipulation, measurement and control of variables

Variables are factors that change or can be changed. In experiments these are the independent and dependent variables as well as any extraneous factors that are or are not controlled. In correlations there are two measured variables (see Section 1.5).

Experiments look for changes or differences in the dependent variable (DV) between two or more levels of the independent variable (IV), which are set up by the experimenter. It is important that the IV is clearly defined, or **operationalised**, so that the manipulation of the conditions represents the intended differences. Consider

KEY TERM

operationalisation: the definition of variables so that they can be accurately manipulated, measured or quantified and replicated. This includes the IV and DV in experiments and the two measured variables in correlations.

RESEARCH METHODS IN PRACTICE

Dr Blot is thinking about buying new chairs for her classroom. Her **aim** is to explore whether hard or soft chairs help her students to work better (Figure 1.11). She wonders whether to predict a **non-directional (two-tailed) hypothesis**: *There is a difference in work rate of students sitting on comfortable and uncomfortable chairs.* Another psychology teacher says that students respond well to other comforts like access to a drinks machine or snack bar, and the soft chairs might make them happier, so they work harder. Dr Blot rewrites her prediction as a **directional (one-tailed) hypothesis**, saying: *Students on comfortable chairs will have a higher work rate than ones sitting on uncomfortable chairs.* A third teacher is not convinced and suggests that if the students are too comfortable they will become sleepy and lazy, so work less. The hypothesis would then be: *Students on comfortable chairs will have a lower work rate than ones sitting on uncomfortable chairs.* Her **null hypothesis** would be: *Any difference in work rate of students sitting on comfortable and uncomfortable chairs is due to chance.*

Now imagine a study which aims to look for a link between sleep and emotions. A non-directional hypothesis might be: *There will be a correlation between amount of sleep and emotional reactivity.* The possible directional hypothesis could say: *There will be a positive correlation between amount of sleep and how emotional someone is.* Remember *not* to say that one factor causes the other to change. The null hypothesis here would be: *Any relationship between amount of sleep and emotional reactivity is due to chance.*

1.11 Would you work harder in lessons if you had more comfortable classroom chairs?

a study testing the effect of age on susceptibility to false memories. The IV would be age, with, for example, 'young', 'middle-aged' and 'old' groups. It is important to know how old the people in the groups are; this is operationalisation. You might operationalise 'young' as under 20 years old, 'middle aged' as 40–50 years old and 'old' as over 70. The DV must also be operationalised, so it can be measured effectively. We could operationalise the DV by counting the number of details 'remembered' about the false memory or how convinced the participants were that it was true.

Controlling variables and standardising procedures

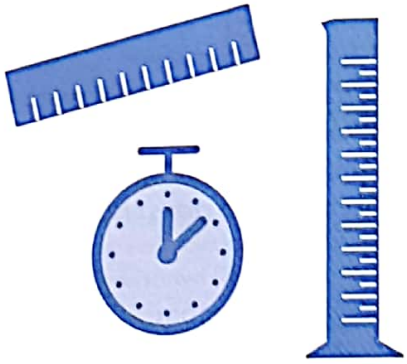
Controlling of variables

Psychologists need to control variables in their studies in order to be more certain about their findings. In particular, in experiments, it is important to control any **extraneous variables** that might have a consistent effect. These are called **confounding variables** as they confound, i.e. confuse, the results. Confounding variables can either work against the effect of the IV or increase the apparent effect of the IV because they act on the DV selectively in one level of the IV. These variables are the most important to control. Other extraneous variables, which have a random effect on the DV across all levels of the IV, are not so problematic. The difficulty is to identify which variables it will be important to control before the experiment starts. This is one function of a **pilot study**, a preliminary test of the procedures of a study. However, if important extraneous variables are not identified in advance, they will become uncontrolled variables, which will affect the results, making them difficult to interpret because it will be hard to separate the effects of the IV from those of the other factors that may have influenced the DV systematically.

Consider Dr Blot's study of students and chairs (Section 1.6). Perhaps Dr Blot compares one class in a room with the new (soft) chairs and another class in a different room with the old (hard) chairs. If the room containing the new chairs happens to have better lighting, Dr Blot may find that the students in the 'soft chairs' condition perform better. However, this may be due to the confounding variable of brighter lighting rather than the comfort level of the chairs. This is an example of a **situational variable**, because lighting is an aspect of the environment. Another possible extraneous variable is how hard working the individual students are. We might expect normally hard-working students to be randomly distributed among the different classes, in which case this variable is not a problem. However, suppose that all the students in the 'soft chairs' class do arts and humanities subjects and all the students in the 'hard chairs' class do maths and sciences. If Dr Blot happens to use a test of data analysis as her measure of the DV, she might find that the students in the 'hard chairs' level of the IV perform better. This would suggest that the soft chairs make students perform worse but could in fact be due to the extraneous variable of subject groups. This is an example of a participant variable, because the difference has been caused by a feature of the individuals, i.e. their ability in maths.

Standardisation

Controls make sure that the levels of the IV represent what they are supposed to, i.e. that the differences between them are going to create the intended situations to test the hypothesis. This helps to ensure validity (see Section 1.11). It is also important that every participant is treated in the same way. This is the process of standardisation. One way that this is achieved is by having standardised instructions, that give the same advice to every person in the study. Imagine a questionnaire testing attitudes to helping behaviour. All participants would need to have the same advice about how to fill it in, so that any effects of social desirability – the influence of needing to give answers that were acceptable to society – were equally likely.



1.12 Scientific instruments are likely to produce objective, reliable data

The procedure itself also needs to be standardised. This involves having equipment or tests that are consistent, i.e. that measure the same variable every time and always do so in the same way. Consider the questionnaire about attitudes to helping again. All the questions should focus on the same aspect of behaviour, i.e. helping, rather than some looking at a different but possibly

KEY TERM

situational variable: a confounding variable caused by an aspect of the environment, e.g. the amount of light or noise.

control: a way to keep a potential extraneous variable constant, e.g. between levels of the IV, to ensure measured differences in the DV are likely to be due to the IV, raising validity.

related factor, such as being friendly or happy. In laboratory experiments, standardisation is easier than in other studies, as equipment is likely to be consistent, for example stopwatches or brain scans. However, some of these measures, such as brain scans, may need to be interpreted and this must also be done in a standardised way (Figure 1.12).

Applying your knowledge of variables and controls to novel research situations

In experiments it is important to be able to decide how to operationalise the IV to produce the different conditions (to achieve validity) and to measure the DV in a consistent (reliable) way and to be able to justify these choices. You will also need to be able to decide what controls it is appropriate to use and to suggest how these can be implemented.

When writing hypotheses, you should ideally operationalise the variables you are referring to. For example, in the hypothesis 'Students using revision apps will learn better than students using mind maps', we do not know how 'better learning' will be measured, or which apps are being used because the variables are not operationalised. To be complete, the hypothesis needs more detail, such as 'Students using the Gojimo revision app will gain better test marks than students using mind maps'. Similarly, the hypothesis 'There will be a correlation between amount of sleep and emotional reactivity' does not operationalise either variable. This could be improved by saying 'There will be a correlation between the number of hours a person sleeps for and their emotional reactivity indicated by how loudly they cry during a sad film'.

RESEARCH METHODS IN PRACTICE

In Dr Blot's experiment about classroom chairs (see Section 1.6), the IV of hard and soft chairs must be operationalised. The text also referred to them as 'comfortable' and 'uncomfortable' chairs, but this still does not make clear what is meant by 'hard' and 'soft'. This could be done by saying *'chairs with wooden/plastic seats'* and *'chairs with padded seats'*. **Operationalisation** of the DV is also needed. The text referred to working 'better' and 'harder' but this is also incomplete. We need to expand on the idea of *work rate*, which was also used. This might be measured by counting the number of pieces of homework handed in late, or the time spent doing extra work. Either of these would indicate the amount of work being done. There are many **extraneous variables** that could be important in this study, for example, some of the students might work harder anyway or the rate of work might vary with the weather. If students worked harder on sunny days, this would be a **situational variable**. The important variables to control are those that could confound the results. For example, if there was a choice of chairs, the students who chose to sit on comfy ones might be the laziest. If left as an **uncontrolled variable**, this could alter the results by making it look as if soft chairs made students work less.

In the description of the correlation on sleep and emotions (Section 1.6), the two measured variables were the *'amount of sleep'* and *'how emotional someone is'* or their *'emotional reactivity'*. It is important to operationalise variables in correlations too. To operationalise these variables we could measure *the amount of time* spent sleeping and ask the participant to fill in a questionnaire about their feelings to measure their emotions.

Reflections: Look at the Research methods in practice box.

For the experiment:

- Suggest **one other** way the IV could be operationalised.
- Suggest **one other** way the DV could be operationalised.
- Would a possible difference between how lazy students were be a *situational variable* or a *participant variable*?
- Suggest **one other** possible *extraneous variable*.

For the correlation:

- Suggest **one other** way the variable of 'emotions' could be operationalised.

1.8 Sampling of participants

A **population** is a group of people (or animals) with one or more characteristics in common. For example, the population of a country is all the people who live there, the population of internet users is everyone who can access the internet. A population could also be people who share a particular interest, such as 'all football supporters' or who have a particular feature, for example 'all left-handed people'. The **sample** is the group of people who participate in a study. They are taken from a population and should ideally be representative of that group so that the findings will be representative. Details about the sample, such as age, ethnicity and gender, are important in most investigations because these features affect many psychological differences.

Other characteristics of the sample, such as socio-economic status, education, employment, geographical location or occupation, may also be relevant. The size of the sample also matters. Small samples are less reliable and are likely to be less representative. The different **sampling techniques** described below produce samples which differ in terms of how well they represent the population. The extent to which they are representative of the population determines how effectively generalisations can be made.

Opportunity sampling

Studies are often conducted with the people who are around at the time. Selecting participants in this way is called **opportunity sampling**. An opportunity sample is unlikely to represent the population fairly because readily available people will tend to be alike so they are unlikely to include the variety that exists. For example, many studies are conducted using university students as they are convenient for the researchers. However, this means that the sample will be predominantly young, with a better than average education. This means that the results may not reflect the scores that people of different ages or educational opportunities might produce. Despite this potential problem, opportunity sampling is the most common method, even for professional psychologists, as for many investigations the results are unlikely to be affected by age or education.

Volunteer (self-selected) sampling

Rather than the researcher choosing individuals to ask, they may invite people to volunteer to take part in their study. They might put up an advertisement, make an announcement or post a request on the internet. In this way the people who respond and become participants choose to do so, i.e. are volunteers, so are described as a **volunteer sample** (Figure 1.13). As the individuals are self-selected, that is they choose whether to join in, this sampling technique is unlikely to be representative of the population. Volunteers may have more free time than average and, apart from being willing, often have other characteristics in common, such as being better educated. Nevertheless, it is a useful technique when looking for participants who are unusual in some way, for example in Baron-Cohen et al.'s study, where people on the autistic spectrum were needed.

KEY TERMS

population: the group, sharing one or more characteristics, from which a sample is drawn.

sample: the group of people selected to represent the population in a study.

sampling technique: the method used to obtain the participants for a study from the population.

opportunity sample: participants are chosen because they are available, e.g. university students are selected because they are present at the university where the research is taking place.

volunteer (self-selected) sample: participants are invited to participate, e.g. through advertisements via email or notices. Those who reply become the sample.

random sample: all members of the population (i.e. possible participants) are allocated numbers and a fixed amount of these are selected in a unbiased way, e.g. by taking numbers from a hat.

Public Announcement

WE WILL PAY YOU $4.00 FOR ONE HOUR OF YOUR TIME

Persons Needed for a Study of Memory

•We will pay five hundred New Haven men to help us complete a scientific study of memory and learning. The study is being done at Yale University.
•Each person who participates will be paid $4.00 (plus 50c carfare) for approximately 1 hour's time. We need you for only one hour: there are no further obligations. You may choose the time you would like to come (evenings, weekdays, or weekends).

•No special training, education, or experience is needed. We want:

Factory workers	Businessmen	Construction workers
City employees	Clerks	Salespeople
Laborers	Professional people	White-collar workers
Barbers	Telephone workers	Others

All persons must be between the ages of 20 and 50. High school and college students cannot be used.
•If you meet these qualifications, fill out the coupon below and mail it now to Professor Stanley Milgram, Department of Psychology, Yale University, New Haven. You will be notified later of the specific time and place of the study. We reserve the right to decline any application.
•You will be paid $4.00 (plus 50c carfare) as soon as you arrive at the laboratory.

TO:
PROF. STANLEY MILGRAM, DEPARTMENT OF PSYCHOLOGY, YALE UNIVERSITY, NEW HAVEN, CONN. I want to take part in this study of memory and learning. I am between the ages of 20 and 50. I will be paid $4.00 (plus 50c carfare) if I participate.

NAME (Please Print).....

ADDRESS .....

TELEPHONE NO. ..... Best time to call you .....

AGE..... OCCUPATION..... SEX.....
CAN YOU COME:

WEEKDAYS ..... EVENINGS .....WEEKENDS.....

1.13 Would you respond to this advert?

Random sampling

Opportunity and volunteer samples may be biased – they will probably contain very similar people so are unlikely to include the spread of characteristics in the population. In **random sampling** each person in the population has an equal chance of being chosen so the sample is much

more likely to be representative. Imagine you are looking for a sample of students at your school and you put an advert for volunteers on the library notice board. Students who never go to the library cannot be included so your sample might be biased towards those who work the hardest. Similarly, if you took an opportunity sample from the common room, it would only include students who are relaxing. Now your sample might be biased towards the *least* hard working. To obtain a representative sample you could instead use a numbered list of all students and use a random number generator to choose the participants. This would be a random sample as any individual is equally likely to be chosen. If the population is small, such as all the members of your class, you can simply give each person a number, put pieces of paper with each number on in a hat, and draw out numbers until there are enough for the sample.

Applying your knowledge of sampling techniques to novel research situations

An early step in any research is to obtain an appropriate sample. The extent to which generalisations can be made from research depends in part on how representative the sample is. It is therefore important to get the best possible sample. However, practical constraints prevent researchers from using random samples most of the time and, for many psychological phenomena, it is reasonable to believe that processes happen in a fairly universal way so some sample bias is unproblematic. However, to assume there are no differences in emotional responses, cognitive processing or behaviour between populations would be misleading. Indeed, the psychology of individual differences, of developmental psychology and cross-cultural research are three areas specifically devoted to the study of such differences. It is therefore important that you can recognise limitations in the sampling technique used. This means that you should be able to identify possible differences between individuals or groups that might matter for the particular phenomena being explored in a study. Imagine two researchers at different universities are both studying obedience and both want samples from people nearby but not students. One university is near a police college and the other is next to a hospital and the researchers both obtain opportunity samples with the same age and gender spreads from these workplaces. Even though the samples are similar in age and gender, the difference in occupations may mean that the results of their studies will be different – because police officers tend to be more obedient than nurses.

You will need to be able to explain how you would use each sampling technique and to explain reasons for choosing each technique. For example, why it might be difficult to use a technique in practice or why generalisations could or could not be made from the sample obtained.

	Sampling technique		
	Opportunity sampling	**Volunteer (self-selected) sampling**	**Random sampling**
Strengths	Quicker and easier than other methods as the participants are readily available	Relatively easy because the participants come to the researcher. They are also likely to be committed, e.g. willing to return for repeat testing	Likely to be representative as all types of people in the population are equally likely to be chosen
Weaknesses	Likely to be non-representative as the variety of people available is likely to be limited, so they will tend to be similar and the sample could therefore be biased	Likely to be non-representative as people who respond to requests may be similar, e.g. all have free time	In reality everyone may not be equally likely to be chosen, e.g. if they cannot be accessed (if the original list is incomplete) or if mainly one type of participant, e.g. girls, happen to be selected. This is particularly important if the sample is small

Table 1.3 Strengths and weaknesses of sampling techniques

RESEARCH METHODS IN PRACTICE

Your teacher has asked you to do a cognitive psychology investigation. You want it to work well so you want a **sample** of 30 people from which you can generalise. You think it would be easiest to take an **opportunity sample** from your classes at the college but you realise this might produce a biased sample as all your subjects are sciences. This might matter as the investigation is about using logic to solve problems and you think that science students might be especially good at the task. If so, their results might not be representative of the college **population** as a whole. A friend suggests making an announcement in the canteen asking for students studying all different subjects. This **volunteer sampling** technique might be better but there would be no guarantee of getting a range of people. You decide that the best sampling technique would be to choose individuals at random from a list of all the students in the college. You hope that this would mean you would be equally likely to get students taking each subject. You decide to use this **random sampling** technique based on the students' examination candidate numbers. You enter the candidate numbers into a random number generator and use the first 30 numbers that are generated.

Reflections: *Look at the Research methods in practice box above. What problems would the following situations lead to?*

- *You revert to your idea of an opportunity sample but use your neighbours, who are mainly retired people.*
- *You followed your friend's advice about volunteer sampling but lots of the younger students were in detention at lunchtime that day.*

Read the examples below. Which sampling technique is being used in each situation?

- *Professor Smudge is doing some internet research and is recruiting participants by posting on Facebook asking for people to help with her study.*
- *Dr Splash is investigating the effects of ageing and is asking all the residents at two local care homes for their help.*
- *Dr Blot pulls student numbers on the college register from a hat to select a sample for a new study on homework and part time jobs.*

1.9 Data and data analysis

Psychologists, like all scientists, often produce numerical results from their investigations. These results are called the 'raw data'. As it is difficult to interpret large amounts of figures, the results are often simplified mathematically and represented visually on graphs. We will discuss a range of methods in this section. Note that you are not required to perform calculations in examinations but you could be asked to count up scores, find the mode, median and range of a data set, make simple comparisons and interpret data from tables or graphs.

Types of data

As you may know from the core studies or from earlier parts of this chapter, psychologists use a variety of different research methods. These methods can produce a range of different types of data. The main types are discussed below.

Quantitative and qualitative data

When psychologists collect data they can collect either numerical results, called **quantitative data**, or **qualitative data**, which is detailed and descriptive. Quantitative data indicates the *quantity* of a psychological measure, such as the amount or strength of a response and tends to be measured on scales, such as time, or as numerical scores on tests such as IQ or personality. Quantitative data is associated with experiments and correlations which use numerical scales but it is also possible to obtain quantitative

KEY TERMS

quantitative data: numerical results about the *quantity* of a psychological measure such as pulse rate or a score on an intelligence test.

qualitative data: descriptive, in-depth results indicating the *quality* of a psychological characteristic, such as responses to open questions in self-reports or case studies and detailed observations.

	Data type	
	Quantitative	**Qualitative**
Strengths	Typically uses objective measures Scales or questions often very reliable Data can be analysed using measures of central tendency and spread making it easy to compare	Data is often valid as participants can express themselves exactly rather than being limited by fixed choices Important but unusual responses are less likely to be ignored because of averaging
Weaknesses	Data collection method often limits responses so the data are less valid, e.g. if the participant wants to give a response that is not available	Data collection is often relatively subjective so findings may be invalid as data recording or interpretation may be biased by the researcher's opinions or feelings Detailed data from one or a few individuals may not generalise to the majority

Table 1.4 Quantitative versus qualitative data

data from observations, questionnaires or interviews. For example, a record of the number of times a behaviour is seen or the total of responses to a closed question in an interview would be quantitative data. The sources of quantitative data are typically highly objective, as the scales or questions used need little if any interpretation, making them high in validity. In addition, the measures used are generally highly reliable, as the measures are fixed quantities.

Qualitative data indicates the *quality* of a psychological characteristic. Such data is more in-depth than quantitative data and includes detailed observer accounts and responses to open questions in questionnaires, interviews or case studies. Although there is a risk of subjectivity in the interpretation of such data by the researcher, qualitative data may be more representative as the participant can express themselves fully, so in some senses qualitative data can also be valid.

RESEARCH METHODS IN PRACTICE

Dr Splash and Professor Smudge disagree over the best way to collect data about people's emotions when they are put in stressful situations. Dr Splash wants to collect **quantitative data** by measuring their pulse rate and give them scales of very / quite / not at all to rate how stressed they feel. Professor Smudge thinks it would be better to collect **qualitative data** by interviewing each participant and getting them to describe their feelings. She plans to ask questions such as *'How do you feel when you meet an important person for the first time?'* and *'Describe how you felt the last time you were late'.*

Reflections: Look at the Research methods in practice box above. What are the advantages and disadvantages of each suggested data collection method?

Are the questions suggested by Professor Smudge open or closed questions?

Using ethical ideas only, suggest two more questions that could be used in the *interview*, **one** *open question* and **one** *closed question*. State which type of question would collect *qualitative data* and which would collect *quantitative data*.

Data analysis

This section explores different ways that mathematics can be used to simplify and understand the data produced by studies in psychology.

Measures of central tendency

A set of quantitative results can be summarised to one number that represents the 'middle' or typical score, called a **measure of central tendency** or 'average'. There are three different measures of central tendency: the **mode**, **median** and **mean**.

KEY TERMS

measure of central tendency: a mathematical way to find the typical or average score from a data set, using the mode, median or mean.

mode: the measure of central tendency that identifies the most frequent score(s) in a data set.

median: the measure of central tendency that identifies the middle score of a data set which is in rank order (smallest to largest). If there are two numbers in the middle they are added together and divided by two.

mean: the measure of central tendency calculated by adding up all the scores and dividing by the number of scores in the data set.

The mode

The mode is the most frequent score in a data set. It can be used with numerical data (such as scores on a test) and also with data sets that are choices that can be counted (such as written responses to the question 'What is your favourite subject: maths, English or psychology?'). If two (or more) values are equally common there will be two (or more) modes. For example, the faces in Figure 1.14 could be used in a test to compare people on the autistic spectrum with a control group. The participants could be asked 'Which face looks the happiest?' The face which was chosen as the happiest by each group would be the mode for that group.

1.14 A facial expression test

A self-report in a school produced the data set in Table 1.5. The mode for subject choice is 'Psychology', because more people said this subject was their favourite – ten compared to four and six for the other subjects.

	Subject		
	Maths	English	Psychology
Number of people	4	6	10

Table 1.5 Number of people choosing each subject as their favourite

Another question asked 'On which day of the week do you do most homework?' (see Table 1.6). The responses from girls and boys were compared.

In Table 1.6 more boys have 'Sunday' as their main homework night, so this is the mode for boys. Girls have that 'Friday' and 'Sunday' are their main homework night but these two categories are the same, so there are two modes, each containing six girls. We could also combine the totals to work out an overall mode. Adding together totals for each day shows that for all students, Sunday is most popular homework night, with 17 students in total.

The median

Unlike the mode, the median cannot be used with data in discrete (separate) categories, it is only used with numeric data on a linear scale (i.e. points in a sequence). To find the median, all the scores in the data set are put in a list from smallest to largest (ranked). The middle one in the list is the median. If there are an even number of participants, so there are two numbers in the middle, these are added together and divided by 2 to find the median.

Another question in the school survey asked participants to rate how hard they thought they worked, from 1 to 10. Students in their AS and A Level years were asked.

AS student data:

8, 6, 9, 1, 5, 6, 2, 7, 3, 6, 9, 8, 5, 6, 3, 8, 5, 10, 2, 3

A Level student data:

7, 9, 6, 7, 9, 7, 10, 10, 7, 10, 9, 4, 9, 6, 10, 10, 7, 9, 7, 7

Putting these data into order for the two groups separately:

AS students:

1, 2, 2, 3, 3, 3, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 8, 9, 9, 10

6 + 6 = 12, 12/2 = 6 so the median = 6

A Level students:

4, 6, 6, 7, 7, 7, 7, 7, 7, 7, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10

7 + 9 = 16, 16/2 = 8 so the median = 8

	Night of the week						
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Number of boys	0	1	2	3	3	0	11
Number of girls	1	1	2	4	6	0	6

Table 1.6 Main homework night

The median for the A Level group, of 8, is higher than the median for AS students, which is 6. This suggests that A Level students believe they are working harder than AS students do.

The mean

The mean is the measure of central tendency that we usually call the 'average'. It can only be used with numerical data from linear scales. The mean is worked out by adding up all the scores in the data set and dividing by the total number of scores (including any that were zeros). It is the most informative measure of central tendency because it takes every score into account.

Looking back at the data used in the section on the median, mean could be calculated instead. There were 20 students in each group. For the AS students, the calculation is therefore all the scores added together, then divided by 20, i.e. 112/20 = 5.6, so the mean is 5.6. For the A Level students, the calculation is again all the scores added together, and divided by 20, i.e. 160/20 = 8, so the mean is 8. Like the median, this too shows that the A Level students believe they are working harder than the AS students do.

Measures of spread

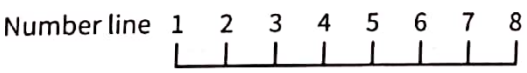
A **measure of spread** is an indicator of how varied the results are within a data set: are they clustered together or widely dispersed? If two data sets are the same size, with the same average, they could still vary in terms of how close the majority of data points were to that average. Differences such as this are described by measures of spread: the **range** and the **standard deviation**.

The range

The range is the simplest measure of spread and is calculated in the following way:

1. Find the largest and smallest value in the data set.
2. Subtract the smallest value from the largest value, then add 1.

You may have learned to calculate the range without adding 1. In psychology we do this because the scales we use measure the gaps between points, not the points themselves. Consider a scale of student happiness from 1 (sad) to 8 (very happy). This can be represented on a line:



If we say someone's happiness is at level of 3, they could be anywhere between 2.5 and 3.5, and someone scoring 6 has a happiness level somewhere between 5.5 and 6.5. So, if 3 and 6 were the lowest and highest scores, the real spread extends to those limits, i.e. from 2.5 to 6.5, so this spread works out as 6.5 – 2.5 = 4. This figure is one bigger than the largest score (6) minus the smallest score (3) = 3.

The range for the two sets of data given in the section on the median would be calculated in the following way:

AS students:

1, 2, 2, 3, 3, 3, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 8, 9, 9, 10

10 – 1 = 9, 9 + 1 = 10 so the range = 10

A Level students:

4, 6, 6, 7, 7, 7, 7, 7, 7, 9, 9, 9, 9, 9, 10, 10, 10,10, 10

10 – 4 = 6, 6 + 1 = 7 so the range = 7

So, not only are the medians and means for these two data sets slightly different, the ranges are different too. This tells us that the diversity of opinion about how hard they are working is greater for AS students than A Level students. We could also say that although most A Level students think they are working very hard, AS student opinion varies from 'not working hard' to 'working very hard'.

One problem with the range is that it does not accurately reflect outliers. That is, it would not be clear from the range whether the most extreme scores, e.g. very large scores, were single odd scores or typical of the data set. Imagine that the least hard-working student in the A Level set had rated themselves as working at level 1 rather than 4. This would make very little difference to the mean (it would be 7.85 instead of 8), but would change the range from 7 to 10 (the same as the range for the AS group).

The standard deviation

In the same way that the mean can tell us more than the mode, a measure of spread called the standard deviation can tell us more than the range. Rather than looking only at the extremes of the data set, the standard deviation (given

KEY TERMS

measure of spread: a mathematical way to describe the variation or dispersion within a data set.

range: the difference between the biggest and smallest values in the data set plus one (a measure of spread).

standard deviation: a calculation of the average difference between each score in the data set and the mean. Bigger values indicate greater variation (a measure of spread).

the name s, SD or σ) considers the difference between each data point and the mean. This is called the *deviation*. These deviations are then squared, added together and the total is divided by the number of scores in the data set, minus 1. The final step is to find the square root. The standard deviation is represented by the formula:

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}}$$

The following symbols are used:

s = standard deviation

x = each score in the data set, i.e. a figure for the variable being measured

$\bar{x}$ = the mean of the data set (called the 'sample mean')

Σ = the Greek letter sigma means 'the sum of', i.e. 'add them all up'

n = the number of scores in the data set

$\sqrt{}$ = the square root.

The deviation, d, is sometimes used in place of the difference between each score in the data set and the mean $(x - \bar{x})$ in the formula above.

As the standard deviation tells us the spread of a group, groups with scores that are more spread out have larger standard deviations, groups with closely clustered scores have smaller standard deviations. When the standard deviations of two groups are similar, this means that they have a similar variation around the mean.

Returning to the data about how hard students believe they are working (from the section on the median) the standard deviations for the two sets of data would be:

AS students: SD = 2.62
A Level students: SD = 1.72

These figures show, as the range did, that there is greater variation in the opinions of the AS students (as the figure of 2.62 for the standard deviation for this group is larger than 1.72, the figure for the A Level group). Note that if the standard deviation is calculated for the A Level group with the opinion score of 4 for one student being replaced with a 1, the standard deviation increases, as this makes the group more varied. However, it still does not become as large as the figure for the AS group. This is different from the effect of the change in this participant's score on the range. The range of the two year groups become the same (they would both be 10), whereas in fact *within* the groups, the AS students are more varied. This is reflected in the standard deviation however, which changes from 1.72 when the student has a score of 4 to 2.16 if they have a score of 1. So, the main advantage of the standard deviation over the range is that it takes every score into account and therefore provides a representation of variation *within* the data set. As it is not just considering the extremes is not distorted by outliers like the range would be.

RESEARCH METHODS IN PRACTICE

A new psychology teacher has tested a group of ten of his students on their recall of the aims, methods, results and conclusions of one of the core studies. The test was marked out of 15. Two students scored 0, and one student scored each of the marks 3–10. The teacher works out the **mode** and is horrified to find that the modal score is zero! A colleague recommends he works out the **median** instead, so he finds the two middle scores in the group (5 and 6), adds them together and divides by 2. This is 5.5, so he feels much better. He wonders if calculating the **mean** will make him feel better still, but it doesn't. When he adds up all the scores and divides by 10 (because he had 10 students), the mean is only 5.2. This is because the median does not consider the value of two zero scores, whereas the mean does. These three figures are different **measures** of **central tendency**. He wants to compare this group to his colleague's students, who scored 0, 1, 1, 2, 2, 2, 3, 4, 4, 5. He calculates two different **measures of spread**. The biggest and smallest values in her set are 0 and 5, so he subtracts 0 from 5 and adds 1, giving a **range** of 6. The **standard deviation** for her group is 1.58.

Reflections:

1 Look back at Figure 1.14. For the data set in the table below, which is the modal response?

	Face A	Face B	Face C	Face D
Number of participants selecting each face as the happiest	0	3	7	1

2 Look back at the data for AS and A Level students in the section on the median. What is the mode for each group?

3 Using the information about the teacher in the Research methods in practice box above answer the following questions:

- What is the mode for the colleague's group?
- Which group does this suggest performed better on the test?
- What is the median for his colleague's group?
- Which group does this suggest performed better on the test?
- The mean for his colleague's group was 2.4. Which group does this suggest performed better on the test?
- What is the range for the teacher's own group?
- When compared to the range for his colleague's group, what does this tell you?
- The standard deviation for his own group was 3.49. Does this tell you that his group was more varied or less varied than hers?

Graphs

A graph is a visual illustration of data. There are many different types of graph and in this section we will consider only bar charts, histograms and scatter graphs, each of which you should be able to name, recognise, draw and use to interpret data. In addition, you will need to be able to recognise, interpret and understand a normal distribution curve.

Bar charts

A **bar chart** is used when the data are in separate categories rather than on a continuous scale. Bar charts are therefore used for the totals of data collected in named categories and for all measures of central tendency (modes, medians or means). The bars on a bar chart must be separate. This is because the *x*-axis represents the distinct groups and not a linear scale. For a bar chart of the results of an experiment, the levels of the IV go along the bottom (on the *x*-axis) and the DV goes on the *y*-axis. To help you to remember which is the *x*-axis and which is the *y*-axis, think 'X is a-*cross*' (Figure 1.15).

1.15 A bar chart showing the number of participants selecting each face as the happiest

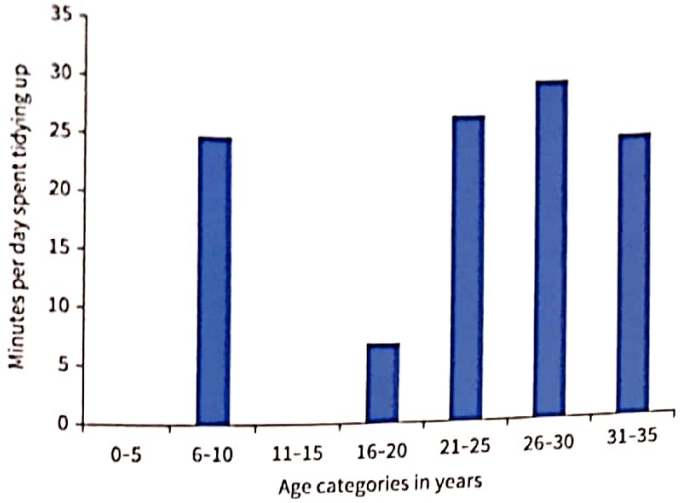
Histograms

Histograms can be used to show the pattern in a whole data set, where this is continuous data, i.e. data measured on a scale rather than in separate categories. A histogram may be used to illustrate the distribution of a set of scores. In this case, the DV is plotted on the *x*-axis (across) and the frequency of each score plotted on the *y*-axis (up the side). The scores along the *x*-axis may be grouped into categories (e.g. if the DV is age, the data may be grouped into 0–5 years, 6–10 years, 11–15 years, etc.). As the scale represented on the *x*-axis is continuous the bars are drawn next to each other, unlike in a bar chart. This means that if there are no scores in a category, a gap must be left to show that the category is empty (see Figure 1.19, which shows 'empty' categories on the *x*-axis).

KEY TERMS

bar chart: a graph used for data in discrete categories and total or average scores. There are gaps between each bar that is plotted on the graph because the columns are not related in a linear way.

histogram: a graph used to illustrate continuous data, e.g. to show the distribution of a set of scores. It has a bar for each score value, or group of scores, along the *x*-axis. The *y*-axis has frequency of each category.



1.16 A histogram of the time spent tidying by people of different ages

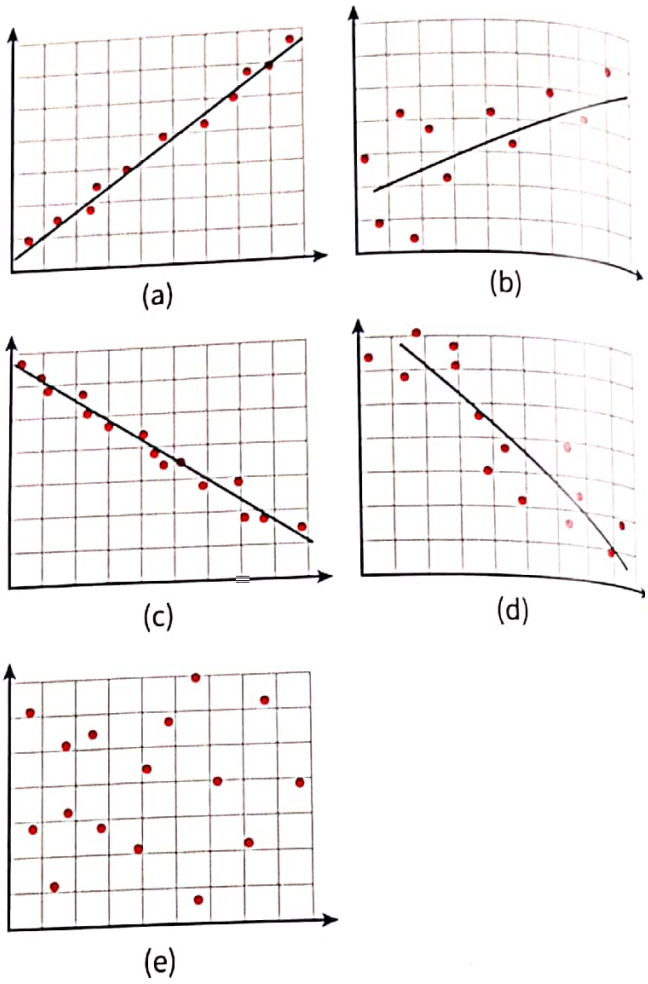
Scatter graphs

Correlations were discussed in Section 1.5. The results from a correlational study are displayed on a scatter graph. To construct a **scatter graph**, a dot is marked at the point where an individual's scores on each variable cross. Sometimes you will see a 'line of best fit' drawn on a scatter graph. The position of this line is calculated and its line is drawn so that it comes close to as many points as possible (see Figure 1.17a-d). In a strong correlation all the data points lie close to the line, but in a weak correlation they are more spread out. Note that you will often see the strength of a correlation described as a number from +1 to –1. Values close to +1 are strong positive correlations and values close to –1 are strong negative correlations. Lower or 'smaller' values (closer to 0) are weaker correlations. Where there is no correlation, the points do not form a clear line (and this has a '*r*' value of 0).

KEY TERMS

scatter graph: a way to display data from a correlational study. Each point on the graph represents the point where one participant's score on each scale for the two measured variables cross.

It is important to remember that you *cannot* draw a causal conclusion from a correlational study. Therefore, scatter graphs such as the ones in Figure 1.17 only tell you that there is a relationship between the variables but not which (if either) of them, is the cause of this link. An experiment could help to find this out.



1.17 Scatter graphs showing (a) strong positive correlation (b) weak positive correlation, (c) strong negative correlati (d) weak negative correlation and (e) no correlation

The normal distribution curve

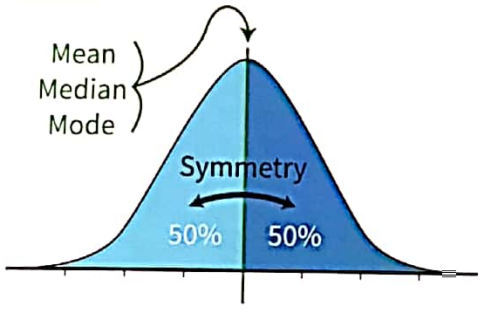
The graph in Figure 1.18 forms a 'bell-shape', which is typica a **normal distribution**. This is a frequency distribution tha

- has the mode, median and mean together in the cent
- has 50% of the scores to the left and 50% to the right the mean
- is symmetrical.

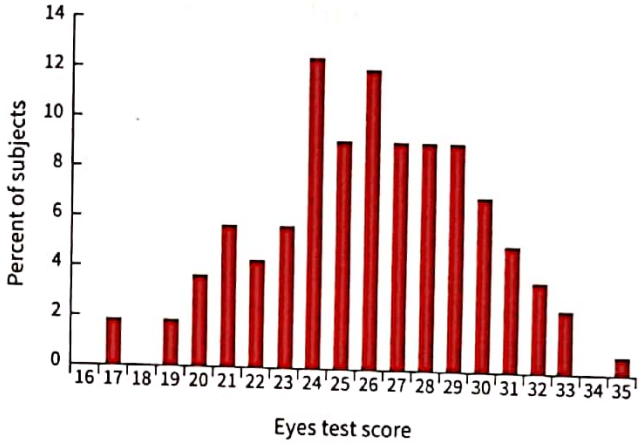
KEY TERM

normal distribution: an even spread of a variable that is symmetrical about the mean, median and mode. The graph showing this distribution is sometimes called a 'bell curve' because of its shape. The graph of the frequency of each score or value rises gradually and symmetrically to a maximum at the point of the mean, median and mode.

A histogram of the distribution of Eyes test scores for non-autistic participants is a normal distribution (see Figure 1.19).



1.18 The normal distribution



1.19 A normal distribution of the Eyes test scores

Reflections: Look at the Research methods in practice box.

- Draw the four graphs suggested, one for each of the groups A–D using the data provided.
- Look at the graph you have drawn for the results from Group C. Do you think this has the general pattern of a normal distribution?

RESEARCH METHODS IN PRACTICE

Groups of students in Dr Blot's class are doing different psychological investigations into altruism. Group A is counting the number of times boys and girls in the class say 'thank you'. They find that on average, the boys said thank you six times and the girls said thank you seven times. They are going to plot a **bar chart** of these averages. Group B is asking everyone the question, 'Which of these situations have you ever helped in: a parent with a pram / a elderly person with their shopping / a child crossing the road / an injured person?' The totals for each group are 14, 7, 16, 3 respectively. They are also going to plot a bar chart. Group C is using an online computer game to test how many seconds it takes each student to respond to another player's request for help. They work out the percentage of students who took 0–5, 6–10, 11–15, 16–20 and 21–25 seconds. The results are 0%, 15%, 45%, 20% and 5%. They plot a **histogram** of their results. Group D asks every person to answer two questions 'How many times have you lent your notes to someone else?' and 'How likely on a scale of 1–10 are you to give away your lunch to somebody who is hungry?' Their results were:

Lend notes	5	3	9	4	2	6	9	7	3	4
Give lunch	7	2	7	10	1	8	8	7	4	5

They are correlating the two sets of scores and will draw a **scatter graph**.

1.10 Ethical considerations

You will encounter **ethical issues** in other chapters. Here we will look at the ethical dilemmas that psychologists face and how they can deal with them effectively.

Ethical issues

As you will have seen from examples of psychological research, investigations using humans or animals have the potential to cause concerns about the welfare of the participants. Such concerns are called ethical issues. Problems may arise through the nature of the study, such as the potential for psychological discomfort caused by a study about stress, or from aspects of the procedure, such as the need to hide the real aim of the study. Ethical issues may also arise from the implications of the research, for example the possibility for results having a negative impact on part of society.

To help psychologists to cope with potential ethical issues that could arise in their research, many countries have an organisation which produces a code of conduct. In addition, research that is being conducted at a university is likely to require approval from the institution's ethical committee. An ethical code provides advice, for example as a set of **ethical guidelines**, that helps psychologists to work in a way that satisfies the primary concern of the welfare of individuals involved in the research as well as the perception of psychology in society. Participants who are deceived or distressed may not want to participate again, may view psychology badly and pass this message on to others, and are less likely to trust the findings of psychological research. These are all outcomes that should be avoided.

Ethical guidelines relating to human participants

The discussion which follows is based on the British Psychological Society Code of Ethics and Conduct (2009), although there are many other similar ethical codes in use throughout the world.

Informed consent

Sometimes it is important in experiments to hide the aims from participants in order to reduce demand characteristics. However, participants have the right to know what will happen in a study so they can give their **informed consent**. The researcher's need to hide the aim makes it hard to get genuine consent. Ideally, full and informed consent should be obtained from participants before the study starts by giving them sufficient information about the procedure to decide whether they want to participate. In some situations it is not even possible to ask for consent. This is often the case in naturalistic observations and field experiments. In such situations, a researcher may ask a group of people similar to those who will become participants whether they would find the study acceptable if they were involved. This is called **presumptive consent** because it allows the researcher to presume that the actual participants would also have agreed to participate if asked.

Especially when participants have not been fully informed, it is important to **debrief** them at the end of the study.

KEY TERMS

ethical issues: problems in research that raise concerns about the welfare of participants (or have the potential for a wider negative impact on society).

ethical guidelines: pieces of advice that guide psychologists to consider the welfare of participants and wider society.

debriefing: giving participants a full explanation of the aims and potential consequences of the study at the end of a study so that they leave in at least as positive a condition as they arrived.

protection of participants: participants should not be exposed to any greater physical or psychological risk than they would expect in their day-to-day life.

deception: participants should not be deliberately misinformed (lied to) about the aim or procedure of the study. If this is unavoidable, the study should be planned to minimise the risk of distress, and participants should be thoroughly debriefed.

Protection (physical and psychological)

A study may have the potential to cause participants psychological harm (e.g. embarrassment or stress) or physical harm (e.g. engaging in risky behaviours or receiving injections). Participants in such studies have the right to be **protected** and should not be exposed to any greater risk than they would be in their normal life. Care should be taken to eliminate such risks (e.g. by screening participants), experienced researchers should be used and studies should be stopped if unexpected risks arise.

Right to withdraw

Participants should be able to leave a study whenever they wish. This is their **right to withdraw** and it must be made clear to participants at the start of the study. Although participants can be offered incentives to join a study, these cannot be taken away if they leave. This

prevents participants thinking that they have to continue. Researchers should not use their position of authority to encourage participants to remain in a study if they want to stop. So in practice, participants may need to be reminded of this right and researchers should follow this guideline even if data will be lost.

Deception

Participants should not be deliberately misinformed, i.e. **deception** should be avoided. When it is essential to deceive participants, they should be told the real aim as soon as possible and be allowed to remove their results if they want to. When participants have been deceived and they know they have been in a study, debriefing (see below) should follow immediately.

Confidentiality

All data should be stored separately from the participants' names and personal information held, and names should never be published unless the individuals have specifically agreed to this. Such information should be stored securely and should not be shared with anyone outside they study. These measures ensure **confidentiality**. The identity of participants should be protected by destroying personal information. However, where it is needed to re-contact participants or to pair up an individual's scores in each condition in a repeated measures design, each participant can be allocated a number which can be used to identify them.

When conducting a case study or field experiment with institutions, confidentiality is still important and identities must be hidden. For example, the names of schools or hospitals should be concealed.

Privacy

1.20 Privacy should still be maintained even if consent cannot be given by participants

Observations, self-reports which ask personal questions and any study which uses personal information risk invading **privacy**. This means that they may enter physical space or emotional territory that the individual would want to keep to themselves. A researcher should make clear to participants their right to ignore questions they do not want to answer. When completing a questionnaire in a laboratory situation, participants should be given an individual space. In observations, people should only be watched in situations where they would expect to be on public display.

The only exception to this is that personally identifiable information can be communicated or published when the participant gives their informed consent for this or in exceptional circumstances when the safety or interests of the individual or others may be at risk.

Debriefing

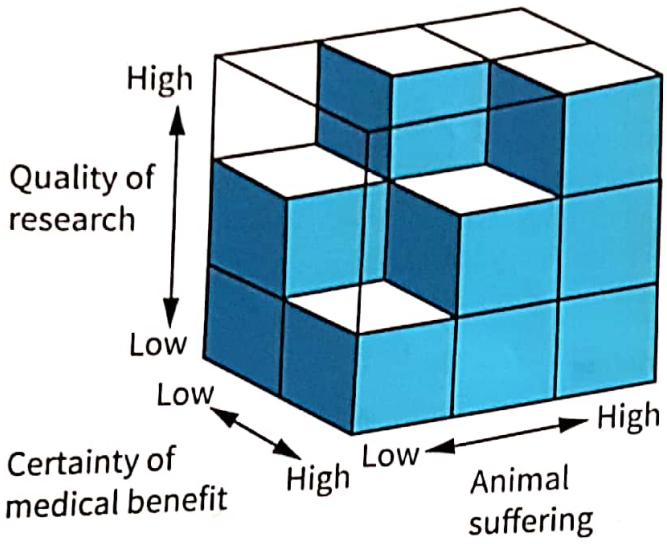
All participants who are aware that they have been in a study should be thanked and given the chance to ask questions. Debriefing participants provides them with an explanation at the end of the study that explains fully the aims of the study and ensures that they do not want to withdraw their data. If participants have been negatively affected by a study the researcher must return them to their previous condition. However, debriefing is not an alternative to designing an ethical study, so it is important to consider all the ways in which a study could cause distress and to minimise them.

Ethical guidelines relating to the use of animals

Animals are used in psychological research for a number of different reasons. Driscoll and Bateson (1988) suggested animals may be: convenient models (e.g. for processes such as learning), a way to carry out procedures that could not be done ethically on humans (e.g. isolation or brain surgery) or be good or interesting examples in their own right (e.g. communication in birds, bats or whales). As a consequence, much psychological research is conducted on animals and therefore their welfare needs protecting.

The discussion which follows is based on the British Psychological Society Guidelines for Psychologists Working with Animals (2012), although there are many other similar ethical codes in use throughout the world. Animals are also often protected by law, but these guidelines specifically consider the effects of research in which animals may be confined, harmed, stressed or in pain, so suffering should be minimised. Veterinary advice should be sought in any case of doubt.

Researchers must aim to ensure that in any research, the means justify the ends, i.e. that the animal suffering caused by the planned experiment is outweighed by the benefits. One way to consider this question is to use Bateson's (1986) cube (see Figure 1.21). When the certainty of benefit (e.g. to humans) is high, the research is good and the suffering is low, the research is worthwhile.



1.21 Bateson's cube (1986)

Replacement

Researchers should consider replacing animal experiments with alternatives, such as videos from previous studies or computer simulations.

Species and strain

The chosen species and strain should be the one least likely to suffer pain or distress. Other relevant factors include whether the animals were bred in captivity, their previous experience of experimentation and the sentience of the species (its ability to think and feel).

Number of animals

Only the minimum number of animals needed to produce valid and reliable results should be used. To minimise the number, pilot studies, reliable measures of the dependent variable, good experimental design and appropriate data analysis should all be used.

Procedures: pain and distress

Research causing death, disease, injury, physiological or psychological distress or discomfort should be avoided. Where possible, designs which improve rather than worsen the animals' experience should be used (e.g. studying the effect of early enrichment on development compared to normal rather than early deprivation). Alternatively, naturally occurring instances may be used (e.g. where stress arises naturally in the animal's environment or lifetime). During research, attention should be paid to the animals' daily care and veterinary needs and any costs to the animals should be justified by the scientific benefit of the work (see Bateson's cube).

Housing

Isolation and crowding can cause animals distress. Caging conditions should depend on the social behaviour of the species (e.g. isolation will be more distressing for social animals than solitary ones). Overcrowding can cause distress and aggression (therefore also physical harm). The level of stress experienced by individuals should also be considered (e.g. the animal's age and gender). Between testing, animals should be housed with enough space to move freely and with sufficient food and water for their health and well-being, both in terms of their biological and ecological needs. However, the artificial environment only needs to recreate the aspects of the natural environment that are important to welfare and survival, e.g. warmth, space for exercise or somewhere to hide. Cage cleaning should balance cleanliness against avoiding stress.

Reward, deprivation and aversive stimuli

In planning studies using deprivation the normal feeding or drinking patterns of the animals should be considered so that their needs can be satisfied (e.g. carnivores eat less frequently than herbivores, young animals need greater access to food and water). The use of preferred food should be considered as an alternative to deprivation (e.g. for rewards in learning studies) and alternatives to aversive stimuli and deprivation should be used where possible.

Anaesthesia, analgesia and euthanasia

Animals should be protected from pain, e.g. relating to surgery using appropriate anaesthesia and analgesia, and killed (euthanised) if suffering lasting pain.

Evaluating studies based on ethical guidelines and applying your knowledge of ethical guidelines to novel research situations

All research with human participants should be ethical. Researchers should, for example, always follow the guideline of confidentiality. Although ethical guidelines should always be followed, sometimes it is necessary to accept some risks to participants in order for a study to achieve its objectives. When this is the case, the researchers

must consider whether the risks can be justified and, if so, how they can be minimised. Ethical issues may arise because of the nature of the topic being studied, such as stress, as this threatens participants' psychological well-being, or they may be at risk of physical harm. Alternatively, ethical issues may arise from the need to use controls which limit participants' knowledge or choices so threaten their privacy, informed consent or right to withdraw. This, in turn, has the potential to cause psychological harm. Although debriefing can help to reduce any harm that has been done, it is not a substitute for designing a more ethical study. Researchers must therefore consider all appropriate guidelines when planning a study, if necessary consulting with colleagues and following the advice of an ethical committee.

You will need to be able to make decisions based on ethical guidelines, and these can be applied to a study you already know about, a novel example that is presented to you or a study that you are asked to design. So, in relation to informed consent, protection (physical and psychological), right to withdraw, deception, confidentiality, privacy and debriefing, you will need to be able to explain why issues have arisen and how and why each guideline has been broken or has been followed.

All research with animals should be ethical. However, in order for any good research to be conducted a compromise must be reached between animal suffering and the good that will come out of the research. In evaluating studies it is important to remember that when the research is planned, decisions must be based on *expected* outcomes, i.e. the potential gains of the study. This is governed by the importance of the work, how certain it is that there will be a valuable benefit and this, in turn, is in part determined by the effectiveness of the research itself. So in deciding whether a piece of research is ethical, we must think about:

- how much the animals suffer (e.g. in terms of pain, deprivation, distress)
- what the positive outcomes might be (e.g. the benefits for people and whether these are worthwhile)
- whether the research is sufficiently well planned to achieve these possible benefits.

You can consider these three elements in terms of Bateson's cube, illustrating each one with examples relating to the particular study. The same ideas can be used whether you are considering a study you already know about, a novel example that is presented to you or a study that you are asked to design. So, in relation to the three points, you will need to be able to explain or justify:

- choices about the animals and their care, such as the species and how they are housed, and the procedures chosen, e.g. the number of animals used and the design of the study in terms of the aspects of feeding, access to companions or pain and distress that arise as a consequence of the study
- why the study is being done, for example in terms of the need for new research or the ways in which humans (or animals) could benefit from the findings
- the strengths of the design of the study, e.g. in terms of controls and ways that good objectivity, validity and reliability have been achieved.

RESEARCH METHODS IN PRACTICE

Professor Smudge is planning an **experiment** on emotions in animals. She wants to see if, like us, they tend to approach things they like and avoid things they don't like. She considered *replacing* the use of animals with videos of animals in the wild responding to different stimuli but decided that she would not know for certain what they were reacting to so the findings might not be valid. She chose to use rats as the *species* is bred for laboratory use and rats can be *housed* alone without distressing them. She decided to use a repeated measures design as it would limit the *number of animals* used compared to an independent measures design. To reduce *pain and distress*, she used food they liked a lot (peanuts) in the 'approach' condition and food they did not like a lot (lettuce) in the 'avoidance' condition. Although the rats were tested before they were given fresh food in their cages, they were not *deprived* of food. As the rats had not been hurt by the procedure, there was no need for *anaesthesia*, *analgesia* or *euthanasia*.

Reflections: Look at the Research methods in practice box above.

- Professor Smudge considered using *aversive stimuli* such as loud noises or electric shocks in the 'avoid' condition but decided this was unethical. Why?
- Suggest **one** problem with Professor Smudge's decision to use a repeated measures design, and explain how this problem might be solved.

1.11 Evaluating research: methodological issues

As well as evaluating research in terms of ethics, it can also be considered in terms of whether it is 'good science', i.e. by looking at methodological issues. There are several key methodological issues that you have encountered elsewhere in the chapter: **reliability**, **validity** and **generalisability**. We will now explore these again, some in a little more depth, and see why they are important in the evaluation of research.

KEY TERMS

reliability: the extent to which a procedure, task or measure is consistent, for example that is would produce the same results with the same people on each occasion.

validity: the extent to which the researcher is testing what they claim to be testing.

generalisability: how widely findings apply, e.g. to other settings and populations.

test–retest: a way to measure the consistency of a test or task. The test is used twice and if the participants' two sets of scores are similar, i.e. correlate well, it has good reliability.

inter-rater reliability: the extent to which two researchers interpreting qualitative responses in a questionnaire (or interview) will produce the same records from the same raw data.

inter-observer reliability: the consistency between two researchers watching the same event, i.e. whether they will produce the same records.

demand characteristics: features of the experimental situation which give away the aims. They can cause participants to try to change their behaviour, e.g. to fit with their beliefs about what is supposed to happen, which reduces the validity of the study.

ecological validity: the extent to which the findings of research in one situation would generalise to other situations. This is influenced by whether the situation (e.g. a laboratory) represents the real world effectively and whether the task is relevant to real life (has *mundane realism*).

Reliability

Whenever research is conducted data is obtained. Researchers must attempt to ensure that the way in which these results are collected is the same each time, otherwise differences could occur (between participants, between conditions in an experiment or between the data obtained by different researchers). Such inconsistencies would be problems of reliability.

The reliability of the measures used to collect data depends on the 'tool' used. A researcher collecting reaction times or pulse rates as data will probably have reliability as the machines used are likely to produce very consistent measures of time or rates. One way to check reliability is to use the **test–retest** procedure. This involves using a measure once, and then using it again in the same situation. If the reliability is high, the same results will be obtained on both occasions, i.e. there will be a high correlation between the two sets of scores. Imagine an experiment on euphoria and anger in which a researcher is not sure whether their questionnaire is a reliable measure of 'happiness'. They use a group of participants and give them the questionnaire on two separate occasions. All the participants would need to be tested at the same time of day and the same day of the week to ensure that their happiness levels were indeed the same. If the 'happiness scale' was reliable, this test–retest procedure would produce a high correlation between the scores on the first and second tests. If the reliability was low, the test would need to be redesigned.

Another reliability problem relates to subjective interpretations of data. For example, a researcher who is using a questionnaire or interview with open questions may find that the same answers could be interpreted in different ways, producing low reliability. If these differences arose between different researchers, this would be an **inter-rater reliability** problem. Similarly if, in an observation, researchers gave different interpretations of the same actions, this would be low **inter-observer reliability**. If the reliability was low, the researchers in either case would need to discuss why the differences arose and find ways to make their interpretations or observations more similar. This can be done by agreeing on operational definitions of the variables being measured and by looking at examples together. These steps would help to make the researchers more objective.

To minimise differences in the way research is conducted that could reduce reliability, standardisation can be used, that is, the procedure is kept the same. This could include instructions, materials and apparatus, although remember that there would be no reason to change many of these. Important aspects of standardisation are those factors which might differ, such as an experimenter's manner towards participants in different levels of the IV,

an interviewer's posture or tone in asking questions or an observer's success at concealing their presence.

Validity

Many factors affect validity, including reliability – a test or task cannot measure what it intends to measure unless it is consistent. Objectivity also affects validity – if a researcher is subjective in their interpretation of data, their findings will not properly reflect the intended measure.

To have **face validity** a test or task must seem to test what it is supposed to. Imagine a test of helping behaviour that involved offering to assist people who were stuck in a bath full of spiders or worms. It might not be a valid test of helping because people who were frightened of spiders or worms would not help, even though they might otherwise be very altruistic. This would be a lack of face validity.

If participants think that they understand the aim of a study, their behaviour or responses are likely to be affected. This would also lower validity. In the design of a study, the researcher should aim to minimise **demand characteristics**, that is, those features which could indicate to the participants what is expected. For example, in a study about false memories, the researcher needs to hide the aim from the participants otherwise they will try to spot which memory is false. They might then try to remember this information particularly well, or might not report it at all if that is what they think the researcher expects.

Another problem for validity is whether the results obtained in one situation will apply to other situations, if not, then the test or task is too specific to be measuring the general phenomenon it was intended for. This is the problem of ecological validity, which applies more widely than considering just whether findings from the laboratory apply to the 'real world'. For example, a test of stress conducted in a laboratory may indeed not reflect the stress experienced in day-to-day life. But equally, a test of stress conducted at home might not reflect the stressful experiences people endure at work or during healthcare procedures. If so, the results may not generalise beyond the situation tested.

The task itself matters too. If a task participants are asked to do is similar to tasks in day-to-day life, it has mundane realism. This is important as the findings are more likely to have high ecological validity if the task is realistic. For example, in an experiment on emotions, responses to dangerous polar bears or dangerous insects could be used. As fewer people would encounter polar bears, responses to insects are likely to have higher mundane realism and therefore higher ecological validity.

Generalisability

Ecological validity contributes to the generalisability of the results. Another factor which affects the ability to generalise is the sample. If the sample is very small, or does not contain a wide range of the different types of people in the population, it is unlikely to be representative. Restricted samples like this are more likely to occur with opportunity or volunteer samples than with a random sample.

Evaluating studies based on methodological issues and applying your knowledge to novel research situations

You will need to be able to make evaluative decisions about methodology in relation to studies you already know about or a novel example of research or a study that you are asked to design yourself. In all of these cases, you need to ask yourself the following questions:

- Are the measures *reliable*? The study will collect data. Does the tool used to collect that data work consistently? Are the researchers using that tool in a consistent way? Is it *objective* or could they be ***subjective*** in their interpretation of the data?
- Is the study *valid*: Does it test what it is supposed to? This may depend on the reality of the task and the *generalisability* of setting (*ecological validity*) or of the sample. Might the participants have been affected by *demand characteristics*?

You will also need to be able to suggest ways to improve the methodology used. You can consider improving the:

- *method* (e.g. a field versus a laboratory experiment or a questionnaire versus an interview)
- *design* (independent measures will have fewer problems with order effects but repeated measures could overcome issues with individual differences)
- *sample* (using opportunity sampling might allow a larger sample to be collected, volunteer sampling could help to find particular types of participants and random sampling would give better generalisability)
- *tool* (measuring the inter-rater reliability or test–retest reliability and changing procedures to make improvements)
- *procedure* (to raise validity by reducing demand characteristics, making the task more realistic, etc.).

RESEARCH METHODS IN PRACTICE

Dr Splash is planning an experiment on obedience. He wants to test whether drivers are more obedient to traffic wardens wearing white clothing or black clothing. He wants the test to be **valid**, so he uses the same male traffic warden wearing different clothing in each condition. He has four observers, one watching cars approaching a junction from each direction. It is important that they are **reliable**, so he gives them operational definitions for the behavioural categories they are to observe:

- *Obedient behaviours*
  - slowing down: visibly reducing speed
  - stopping: coming to a halt before the line on the road.
- *Disobedient behaviours*
  - stopping late: coming to a halt past the line on the road
  - driving on: failing to stop when instructed to do so by the traffic warden.

These categories aimed to be very **objective**. He believes that it is unlikely that the participants (the drivers) will respond to **demand characteristics** as they would not know that they were in an experiment.

Finally, the drivers who stop past the line are given a note by a confederate while they are stationary. This debriefs them and asks if they would be happy to answer questions by telephone. Dr Splash's office number is given for them to call.

Reflections: Look at the Research methods in practice box above.

- Dr Splash is concerned about the generalisability of his findings. He has two ideas for changes to the procedure: conducting the same test in a village rather than a town and using a female traffic warden in both conditions. Explain how each idea would improve generalisability.
- The proposed study has high *ecological validity*. Explain why.
- Dr Splash thinks that of the observations in one of the behavioural categories, visibly reducing speed could be subjective. Explain why this is likely.
- Dr Splash wants to measure the *inter-observer reliability* of his four observers. Explain why this is important.
- In the final part of the study, some participants find out that they have been in a study. Suggest one ethical problem that could arise from this.
- By giving the drivers a number to call, rather than taking their number and calling them, Dr Splash is giving the participants their *right to withdraw*. Why is this important?
- Dr Splash asks the participants who do call him why they stopped and why they stopped over the line. He asks two of his colleagues to interpret the reasons they give but wants to ensure that they have high inter-rater reliability. He gives them a list of possible interpretations including a numerical scale to indicate how strongly the participant felt they may be punished. As both colleagues interpreted the responses from all the available drivers, Dr Splash can correlate the score given to each driver by the two colleagues to see if they are similar. What can he conclude if this produces a strong positive correlation?

SELF-ASSESSMENT QUESTIONS

6 A student is designing an experiment which aims to test whether dogs are more intelligent than cats. He has three dogs and two cats which he plans to use as his sample. To find out which is most intelligent, he is going to hide their food bowl inside a box and time how long it takes the animal to get to the food.
   - **a** Identify and operationalise the independent variable in this experiment.
   - **b** Identify and operationalise the dependent variable in this experiment.
   - **c** Write a non-directional hypothesis for this experiment.
   - **d** Write a null hypothesis for this experiment.
   - **e** Identify and outline the sampling technique used in this study.
   - **f** Which measure of central tendency would be best to find out the average time taken to find the food?
   - **g** A friend suggests that this is not a very valid test of intelligence because it might depend on how well the animal can smell the food. Explain this criticism.

7 A study into sleep obtained participants by placing advertisements in shops near to the university. The participants who responded were a sample of nine females and one male and were mainly retired people. The study was testing a new way to help people to fall asleep, using a recording of bubbling stream. Half the participants were told it would help them to sleep, the others were told it would keep them awake.
   - **a** Identify and outline the sampling technique used in this study.
   - **b** Explain **one** possible problem with generalisability in this study.
   - **c** Describe **one** ethical issue that would arise in this study.
   - **d** How well the participants slept was measured in two ways, by self-report and by how many minutes they stayed asleep for. Which of these measures is more reliable and why?

Summary

Psychologists can use several different **research methods**; **experiments** (**laboratory**, **field** and **natural**), **self-reports** (**questionnaires** and **interviews**), **case studies** (detailed investigations of a single instance, e.g. one person), **observations** and **correlations**.

In experiments there is an **independent variable** (IV), which is manipulated, changed or (in natural experiments) used to create different conditions and a measured **dependent variable** (DV). By imposing **controls**, the experimenter can be more certain that changes in the IV are the cause of changes in the DV. There are three **experimental designs**. In an **independent measures design** there are different participants in each level of the IV, in a **repeated measures design** the same participants are used in all levels of the IV and in a **matched pairs design** the participants are paired up with one member of each pair in each level of the IV. In a repeated measures design **counterbalancing** helps to overcome **order effects** (**fatigue** and **practice effects**) and in an independent measures design **random allocation** helps to overcome the effects of individual differences. In experiments it is important to control variables to raise validity. The most important are **extraneous variables** which could have a confounding effect. If these are left as **uncontrolled variables** they can alter the apparent effect of the IV on the DV. Variables can be described as **participant variables** (due to differences between individuals or between the same individual at different times) or **situational variables** (due to differences in physical setting or the social situation).

In self-reports, different question types can be used, including **open questions** (producing qualitative data) and **closed questions** (producing quantitative data). An interview can be **structured** (fixed), **unstructured** (variable) or **semi-structured**. Observations can be conducted in many different ways, for example **structured** (observing known categories) or **unstructured** (recording any events) and **naturalistic** (observing whatever is happening) or **controlled** (constructing events to observe). The role of the observer may be obvious to the participants (**overt**) or hidden (**covert**) and the observer themselves may be part of the social situations (**participant**) or not (**non-participant**). Correlations look for relationships between two measured variables. They can be **positive** (the two variables increase together) or **negative** (as one variable increases the other decreases) but conclusions cannot be drawn about causal relationships between the variables. All variables, e.g. those in correlations, the IV and DV in experiments and behavioural categories in observations should all be **operationalised**.

Any research begins with an **aim**, which is developed into a testable **hypothesis**. This can be **directional (one-tailed)** or **non-directional (two-tailed)**. This is compared to a **null hypothesis**, which proposes that there is no difference or relationship (or that any pattern in the results has arisen due to chance). To test the hypothesis, a group of participants (the **sample**) is selected from the **population**. This can be done by **opportunity sampling** (choosing people who are available), **random sampling** (selecting participants so that each individual has an equal chance of being chosen) or **volunteer** (self-selecting) **sampling** (inviting participants, e.g. by advertising).

Studies can collect different types of data. **Quantitative data** is numerical and **qualitative data** is descriptive. **Data analysis** of quantitative data includes using various **measures of central tendency** (the **mean**, **median** and **mode**) and **measures of spread** (the **range** and **standard deviation**). Data can be displayed graphically using **bar charts**, **histograms** or **scatter graphs**.

The **normal distribution** is a pattern which can be seen on a frequency histogram which shows that the results have an even (symmetrical) spread around the mean, median and mode.

Research in psychology raises **ethical issues**. Some important issues relate to **informed consent** (knowing about the study and agreeing to do it), **protection of participants** (**physically** and **psychologically**), the **right to withdraw** (being able to leave a study), **deception** (being misled), **confidentiality** (keeping participants' data anonymous), **privacy** (not invading physical or mental space) and **debriefing** (explaining the study to participants afterwards and returning them to their previous state). There are also ethical guidelines relating to the use of animals, including issues relating to the species used, number of animals, the pain and distress they experience, the way they are housed and rewarded or deprived and their suffering (the need for anaesthesia, analgesia and euthanasia).

Two very important **methodological issues** are **validity** and **reliability**. **Ecological validity** relates to how well the findings from one situation, e.g. a laboratory, represent what would happen in other situations. **Subjectivity** threatens validity because it causes researchers to interpret findings from their personal viewpoint, whereas **objectivity** allows researchers to measure variables in ways that are independent of their own perspective. **Demand characteristics** also threaten validity because they inform participants about the aim of the study which can alter their behaviour. Results of studies should be **generalisable**, that is they should apply to other people, situations and times. **Reliability** refers to the consistency of measures. In an experiment it is important to use **standardisation** of procedures to ensure that all participants are treated in the same way. This raises reliability. When researchers interpreting data are consistent, they have good **inter-rater reliability** (e.g. due to practice of operational definitions). **Inter-observer reliability** is the consistency in the records made by observers who are watching the same events. The reliability of a test, e.g. a questionnaire or a task in an experiment can be evaluated using a procedure to measure **test–retest reliability**, by conducting the test twice and correlating the two sets of data.

Exam-style questions

1 A hypothesis in a study says 'Greater emotions will be experienced after an adrenalin injection than after a saline injection'.

a Is this a directional (one-tailed) hypothesis or a non-directional (two-tailed) hypothesis? Include a reason for your answer. **[1 mark]**

b Write a null hypothesis that could be used with the hypothesis given above. **[2 marks]**

2 Declan is conducting a self-report study about attitudes to people with phobias. He cannot decide whether to use a questionnaire or an interview.

a Suggest **one** advantage and **one** disadvantage of using a questionnaire for Declan's study. **[4 marks]**

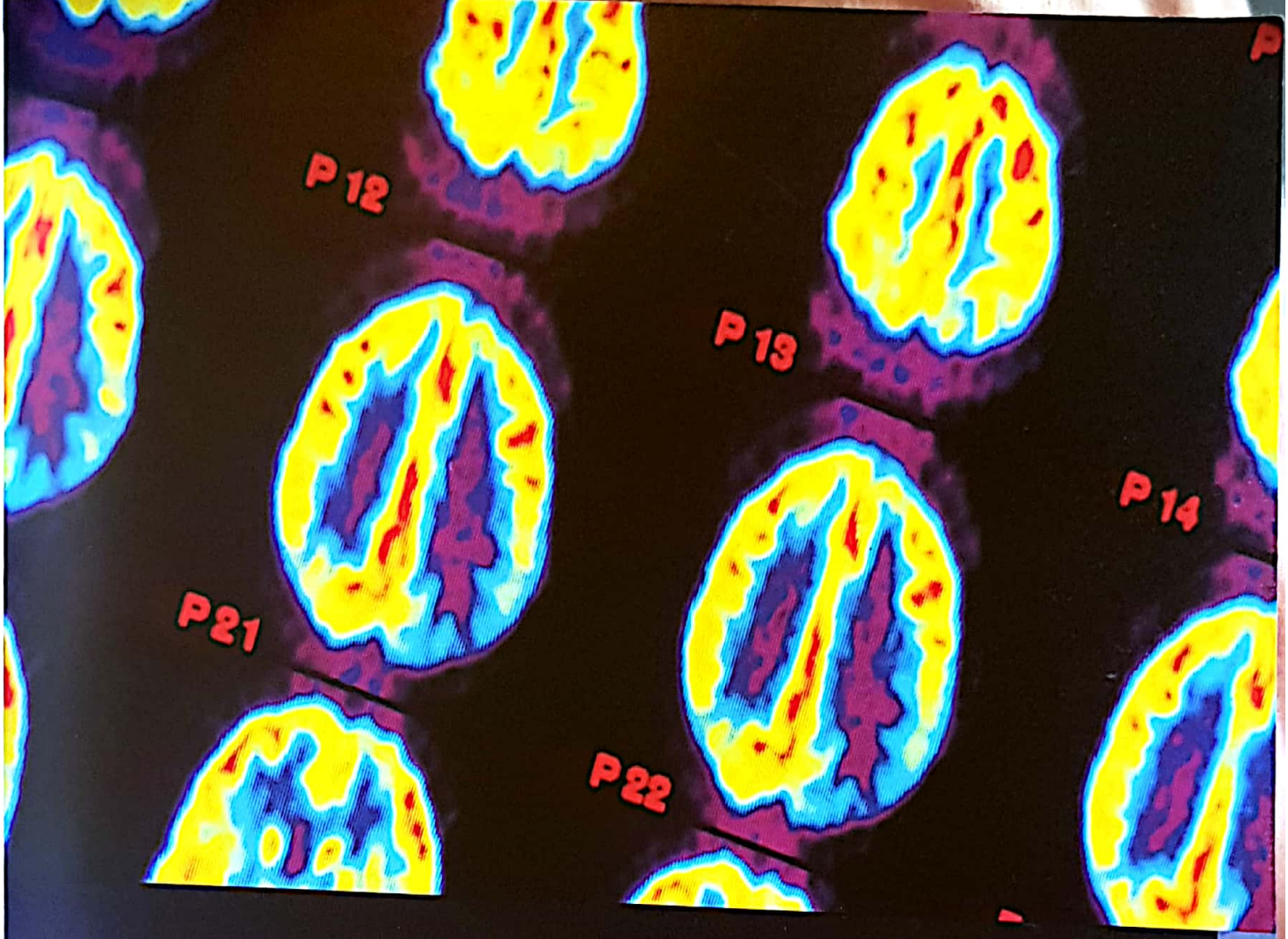
b Write **one** open and **one** closed question that Declan could ask. **[2 marks]**

c Declan is concerned that his interpretation of the responses to questions might not be consistent. Is this mainly a reliability or a validity issue? Explain your answer. **[2 marks]**

3 Mary is planning an experiment to find out whether boys or girls in her school doodle more.

a Describe how Mary could conduct her experiment. **[10 marks]**

b Identify **one** possible weakness / limitation with the procedure you have described in your answer to part (a) and suggest how your study might be done differently to overcome the problem. **[4 marks]**



Chapter 2
The biological approach

Introduction

The aim of this chapter is to introduce you to the biological approach to psychology and to explore three studies from this approach. They are:

- **Canli et al. (2000)** which is a brain scanning study looking at the links between the amygdala and memory for emotional experiences.
- **Dement and Kleitman (1957)** which is a study using a range of methods to investigate the relationship between dream content and eye movements.
- **Schachter and Singer (1962)** which is an experiment that explores the of role of two factors, cognition and physiology, in our experience of emotions.

These three studies illustrate the main assumptions of the biological approach, which are that emotions, behaviour and cognition:

- are controlled by biological systems and processes, such as evolution, genes, the nervous system and hormones
- can be investigated by manipulating and measuring biological responses, such as eye movements, brain activity and pulse rate.

In this chapter there are examples of different aspects of biological factors. In each core study you will learn about the background to the research, the way it was conducted, the findings, conclusions and an evaluation of the study. From this you will be able to see how the studies, and the biological approach in general, illustrate a range of aspects from the research methods topic. In addition, you will discover how these ideas can be used to illustrate a range of issues and debates.

A biological being

What have you been thinking, doing and feeling today? Your answer could range from 'nothing at all' to 'I've run a marathon', 'I've sat an exam' or 'I've cried because I couldn't do my homework'. All of those things are ultimately controlled by your biology. Even if you were doing nothing, your brain was active, that is a biological process of electrical and chemical signals along and between nerve cells (neurons) was happening. The movements when you run are controlled by your brain and messages are sent along the neurons inside your arms and legs. The decisions you make answering exam questions are controlled by your brain too. Emotional responses like crying, even though we cannot control them, are governed by the brain, although hormones are important too. A hormone called adrenalin would be released during the excitement of a race and would help you to run faster. Hormones are often involved in emotional responses too, like being very happy or being sad or angry.

Hormones are released in particular situations. For example, adrenalin would be released if you were running away because you were scared. Adrenalin has effects that would help you to run faster, such as providing extra blood to the muscles. Biological responses like this have evolved because they help us to survive – by running faster we can stay safe. In order to be affected by evolution, aspects of a response, or the physiology that controls it, must be genetically controlled. Imagine a situation in which you were sleeping, and dreaming about jumping out of a window. If you actually did this in your sleep, it would be very risky. However, a system has evolved to protect us. When we are dreaming almost all of the muscles we use for movement (except the ones of our eyes) are paralysed.

Reflections: Think of a behaviour or an emotional response that could have been useful to survival in the early **evolution** of humans. Do you think it could be (partly) controlled biologically, by **genes** and hormones or the nervous system?

2.1 Core study 1: Canli et al. (2000)

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience, 20*, 1–5.

Aim

Canli et al. aimed to show that emotive images will be remembered better than those that have little emotional impact on an individual.

The central questions addressed by this study were whether the amygdala is sensitive to varying degrees of emotional intensity to external stimuli and whether the level of intensity enhances memory for the stimuli.

Background

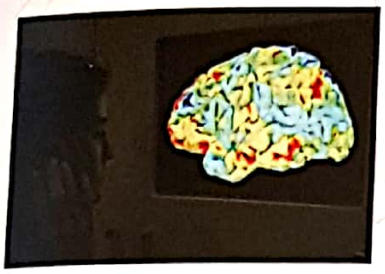
Brain scanning techniques are a huge advance in biological psychology. Psychologists can now study the brains of living people and draw conclusions about the relationship between behaviour and brain structure/activity. There are two basic types of medical scan: functional and structural. Structural scans take detailed pictures of the structure of the brain whereas functional scans are able to show activity levels in different areas of the brain.

Functional magnetic resonance imaging (fMRI) is a neuroimaging procedure using MRI technology that measures brain activity by detecting changes associated with blood flow. In the simplest fMRI study a participant would alternate between periods of completing a specific task and a control or rest state to measure baseline activity. The fMRI data is then analysed to identify brain areas in which the signal changed between the activity and the rest state and it is inferred that these areas were activated by the task.

KEY TERMS

evolution: the process of natural selection of offspring which have inherited characteristics that make them most likely to survive.

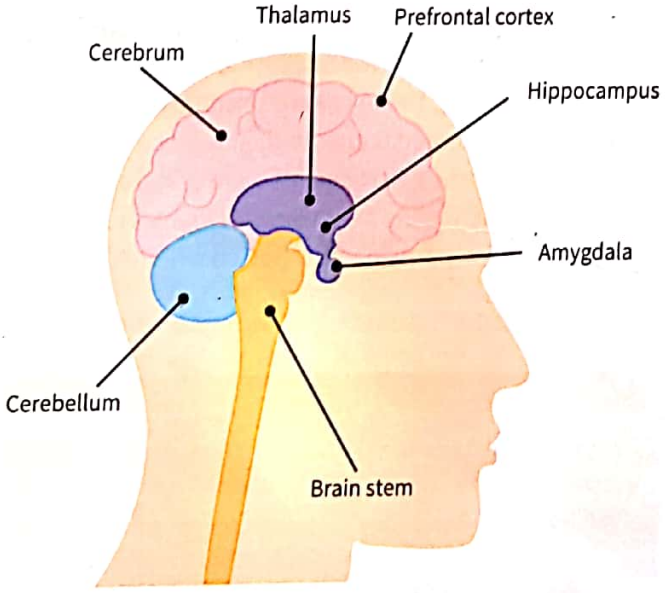
genes: inherited instructions that are passed on from parent to children that control our development and influence some aspects of our thinking, behaviour and emotions, such as our personality and intelligence. One way this can happen is by affecting brain function.



2.1 A computer-generated image using fMRI data showing areas of activity in a human brain

The data from an fMRI scan is used to generate images that can illustrate how the brain is working during different tasks (Figure 2.1). Such a scan allows a living brain to be seen without resorting to surgery. During the scan, patients are placed in a scanner that sends a strong magnetic field through their head. The magnetic field causes the nuclei in hydrogen molecules in the brain to spin in a particular way, and the scanner picks this up. Because hydrogen concentrations vary in different parts of the brain, the scanner is able to create a very detailed picture of the brain.

Over the last few decades, researchers have used fMRI scans to identify areas of the brain that have specific functions (Figure 2.2). Areas that have been shown to have a significant association with emotion and memory are the subcortical areas of the brain, including the amygdala. The amygdala is an almond-shaped set of neurons located deep in the brain's medial temporal lobe and has been shown to play a key role in the processing of emotions such as pleasure, fear and anger. Importantly, the amygdala is also responsible for determining where memories are stored in the brain and which ones are kept.



2.2 Brain map

LaBar and Phelps (1998) suggested that emotional experiences are often better recalled than non-emotional ones and emotional arousal appears to increase the likelihood of memory consolidation during the storage stage of memory (the process of creating a permanent record of the encoded information). Brain imaging studies have shown that amygdala activation correlates with emotional memory in the brain. Previous research by Canli et al. (1999) showed that participants who had a strong amygdala activation in response to a set of emotional stimuli also showed superior memory for those stimuli. However, Canli et al. (2000) suggested that, because an independent measures design was used for these experiments, there could be other explanations for the findings.

The present study used fMRI in a repeated measures, subsequent-memory design to test the predictions that those emotionally intense stimuli that produce greater amygdala activation would be recalled more easily than stimuli that generate less amygdala activation. Participants saw neutral and negative scenes and indicated how they experienced the emotional intensity in each case. A separate fMRI response was recorded in the amygdala for each such emotional experience. Three weeks later, participants' memories for the experiences were assessed to see if those images that generated greater activation of the amygdala were remembered better.

Reflections: If you had the use of a brain scanner what behaviour would you want to locate and why?

Is it an advantage for the human brain to have specialised locations for some tasks or is there a benefit to having tasks distributed across the brain?

Method

Research method and design

This was a laboratory experiment as the environment in which the participants were tested was not comparable to an everyday situation. The use of fMRI scanners limits the realism that can be introduced into a study as they are large machines (see Figure 2.3) and participants have to lie very still while the measurements are being taken.

The independent variable was can be seen as the intensity of the emotional arousal to each of the 96 scenes that were presented to each participant. Participants had to choose from four buttons to indicate emotional arousal on a scale from 0 ('not emotionally intense at all') to 3 ('extremely emotionally intense').

This experiment is an example of a repeated measures design as participants contributed to each of the four conditions depending on their rating of each scene.

IV Level of perceived emotional arousal	Not emotionally intense at all …		… Extremely emotionally intense	
	0	1	2	3
DVs	fMRI measure of amygdala activation			
	Memory of scene			

There were two key measures of the dependent variable. The first was the level of activation of the amygdala measured by fMRI during the first stage of the experiment when the participants were exposed to each of the 96 scenes. During functional scanning, 11 frames were captured per trial, so therefore for each of the 96 scenes there were 11 fMRI measures of neural activity. The second was the measure of memory when participants had to recognise the images three weeks after the initial experiment.

Sample

Ten right-handed healthy female volunteers were scanned. Females were chosen in this study because it was thought that they are more likely to report intense emotional experiences and show more **physiological** reactions to the stimuli.

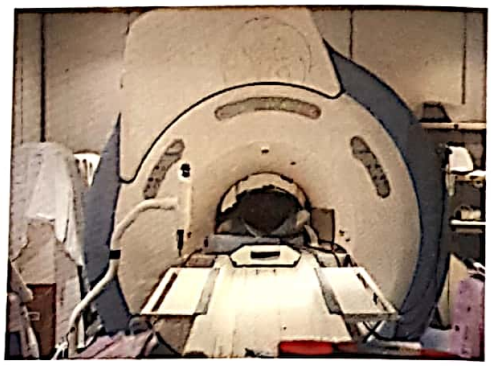
Procedure

During scanning, participants viewed a series of 96 scenes that were presented via an overhead projector and mirror to allow them to see it while in the fMRI scanner. All of the participants had given informed consent to be involved in the study and were aware of the nature of the experiment. The individuals who were operating the fMRI scanner were fully trained and competent in the safety arrangements that should be followed during a medical scan.

All of the 96 scenes were from the 'International Affective Picture System' stimuli set. For the scenes used in this study, average ratings for **valence** ranged from 1.27 (highly negative) to 5.44 (neutral).

The order of scenes were randomised across the participants, with each picture presented for a period of 2.88 seconds. There was then an interval of 12.96 seconds where participants viewed a fixation cross. Participants were instructed to view each picture for the entire time that it was displayed, and after its replacement with a fixation cross they were to indicate their emotional arousal by pressing a button with their right hand. Participants had to choose from four buttons to indicate emotional arousal on a scale from 0 ('not emotionally intense at all') to 3 ('extremely emotionally intense').

To measure activity in the brain, fMRI data were collected by a 1.5 Tesla fMRI scanner (Figure 2.3), which was used to measure blood-oxygen level-dependent contrast. Contrast imaging is a method used in fMRI to observe different areas of the brain which are found to be active at any given time.



2.3 1.5T General Electric Signa MR imager as used in the research

Three weeks after the first stage, participants were tested in an unexpected recognition test in the laboratory. During this task they viewed all of the 96 previous scenes and 48 new scenes (**foils**). The foils were selected to match the previously presented scenes in their valence and arousal characteristics. Participants were asked whether they had seen each scene before and for images judged as previously seen, participants reported whether they

RESEARCH METHODS

Randomising the order of stimuli helps to overcome order effects, so that seeing one type of stimulus doesn't affect the response to stimuli that follow in a systematic way. This means, for example, that always seeing negative scenes could not consistently affect the perception of neutral scenes. Can you remember the two different types of 'order effects'?

KEY TERMS

physiological: to do with the biological processes in the body, for example hormones.

valence: when discussing emotions this refers to the attractiveness (positive valence) or aversiveness (negative valence) of an event, object or situation.

foil: an unknown or unseen object that is used as a control when testing a participant's memory.

remembered with certainty ('remember') or had a less certain feeling of familiarity ('know').

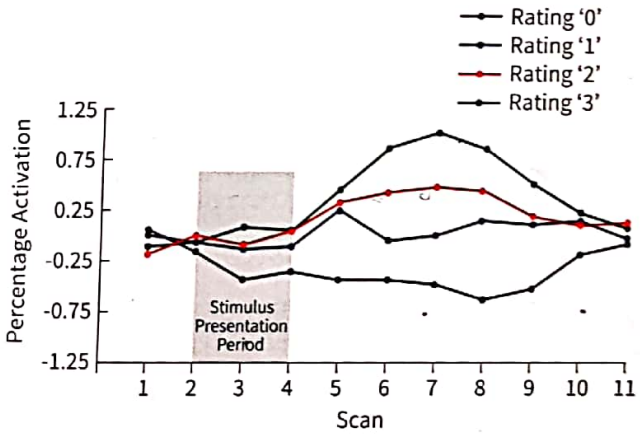
Research questions

Is the amygdala sensitive to varying degrees of individually experienced emotional intensity?

What degree of emotional intensity affects the role of the amygdala in enhancing memory for emotional stimuli?

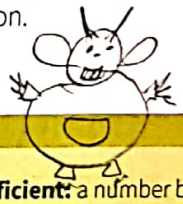
Results

Participants' experience of emotional intensity in the present study correlated well with average ratings of emotional valence and arousal. The average **correlation coefficients** between participants' intensity ratings, on the one hand, and normative valence and arousal, on the other hand, were -0.66 and 0.68, respectively. Therefore, participants' ratings of emotional intensity reflected equally well the valence and arousal characteristics of the stimuli.



2.4 Graph showing average amygdala activation in response to scenes that were rated in emotional intensity from 0 (least intense) to 3 (most intense)

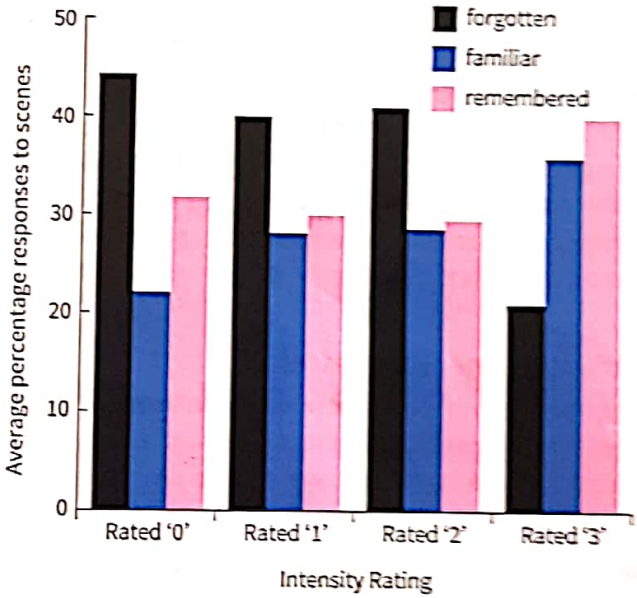
Amygdala activation was significantly correlated with higher ratings of individually experienced emotional intensity (Figure 2.4). This provides evidence that amygdala activation is related to the subjective sense of emotional intensity and that the participants' perceived arousal is associated with amygdala activation.



KEY TERM

correlation coefficient: a number between –1 and 1 which shows the strength of a relationship between two variables with a coefficient of –1 meaning there is a perfect negative correlation and a coefficient of 1 meaning there is a perfect positive correlation.

The follow-up memory task indicated that memory performance was significantly improved for scenes that were rated as highly emotionally intense (i.e. rated 3) than for scenes rated less emotionally intense. Scenes that were rated mild-to-moderate (ratings 0 to 2) had similar distributions of items that were forgotten, familiar, or remembered, whereas scenes that were rated as emotionally highly salient (rated 3) were recalled better, because fewer items were forgotten and more were familiar and remembered (see Figure 2.5).



2.5 Average percentage of scenes forgotten or rated as familiar/remembered at the three-week unexpected recognition task

RESEARCH METHODS

This is a special kind of **bar chart.** Imagine it is really three bar charts, one made of each different coloured set of bars. Each bar chart would be drawn on the same axes. By presenting them all together, it is easier to compare the different types of remembering and forgetting. Why is a bar chart being used here rather than a histogram?

For scenes that were rated highly emotional (rated 3), the degree of left (but not right) amygdala activation predicted whether individual stimuli would be forgotten, appear familiar, or be remembered in a later memory test. Therefore, little amygdala activation when viewing a picture rated as highly emotionally intense was associated with the participant's forgetting the stimulus, but intermediate and high amygdala activation was associated with a participant's later report of familiarity or confident recognition.

Conclusions

Canli et al. found an association between individual experiences of emotional intensity for stimuli with amygdala activation and subsequent memory for these stimuli, suggesting that the more emotionally intense an image is, the more likely it will be remembered. This conclusion provides evidence to explain why people remember emotionally intense experiences well. The level of arousal a person is under could affect the strength of a memory trace. When exposed to an event that causes this arousal, such as a car crash or witnessing a crime, the memory trace will be more robust.

They also found that the amygdala is sensitive to individuals' experienced emotional intensity of visual stimuli with activity in the left amygdala during encoding being predictive of subsequent memory. Canli et al. do comment that some of their findings are correlational, showing an association between the emotional impact on the participant and the subsequent memory for the item.

Strengths and weaknesses

The main method was a **laboratory experiment** as all of the participants were tested in a standardised environment and given the same items to rate in each condition. The procedure was incredibly standardised to the level of the time that each item was presented for and the duration of the interval between presentations. This means that the research has **internal validity** as the researcher can be more confident that there are fewer confounding variables affecting the variables that they are measuring.

The use of an fMRI scanner to measure one of the dependent variables provided the researchers with vast amounts of **quantitative data** relating to the activation of the amygdala, enabling them to carry out statistical analyses such as correlational analysis about the level of activation and subsequent memory of the scene. Although the task of being scanned in an fMRI scanner is hardly **ecologically valid** it would not be possible for the participants to respond to demand characteristics, which increases the validity of the data collected and allows for very sophisticated analysis.

Using fMRI scanners to measure the biological response on the brain of the different images gives an objective finding for each participant as it does not need the research to interpret any results. However, we need to be careful that we do not infer too much from fMRI results as there are still many unknowns about locations of specific behaviours in the brain. Although the method is widespread, there is insufficient knowledge of the physiological basis of the fMRI signal to interpret the data confidently with respect to neural activity and how this maps onto specific behaviours.

Although most fMRI research uses contrast imaging as a method to determine which parts of the brain are most active, because the signals are relative to the individual rather than measured against a baseline, some question the validity in comparisons across individuals. However, in this research, a repeated measures design was used to minimise the impact of this, as participants' scores were compared against themselves.

The sample in this research consists of right-handed females only and therefore introduces participant variables that could distort the outcomes of the research reducing the validity of the research. Females were chosen as the researchers felt that they would have a higher emotional response and might react differently to the scenes that were presented. Therefore, we must be cautious in generalising the results of this to males and to left-handed individuals, as they may respond differently, for example females were chosen rather than males exactly because they were thought to be more emotionally reactive.

SELF-ASSESSMENT QUESTIONS

1 Give **two** examples of how Canli et al.'s study lacks ecological validity.

2 A feature of laboratory experiments is that the researchers have control over the situation and can standardise their procedure. Give **four** examples of controls in this experiment.

3 **a** Describe the sample in Canli et al.'s experiment.

b Evaluate the sample in relation to Canli et al.'s ability to generalise their findings to the general population.

2.2 Core study 2: Dement and Kleitman (1957)

Dement, W., & Kleitman, N. (1957). The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology, 53*, 339–346.

Aim

The aim of this study was to find out more about dreaming. This included three specific research questions:

1 Does dream recall differ between eye movement (**REM**) and quiescent (**nREM**) stages of sleep?

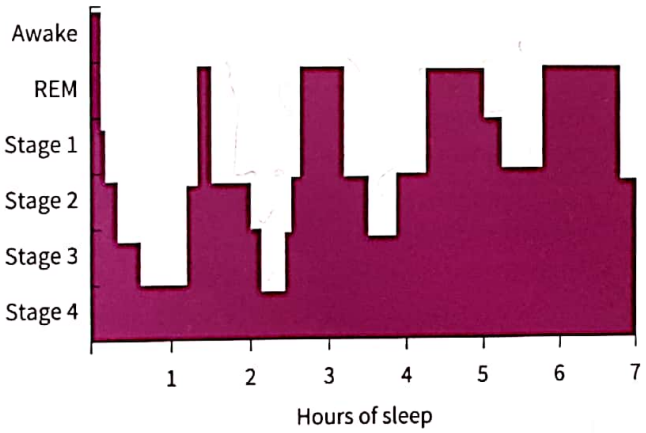
2 Is there a positive correlation between subjective estimates of dream duration and the length of the REM period before waking?

3 Are eye movement patterns related to dream content?

Background

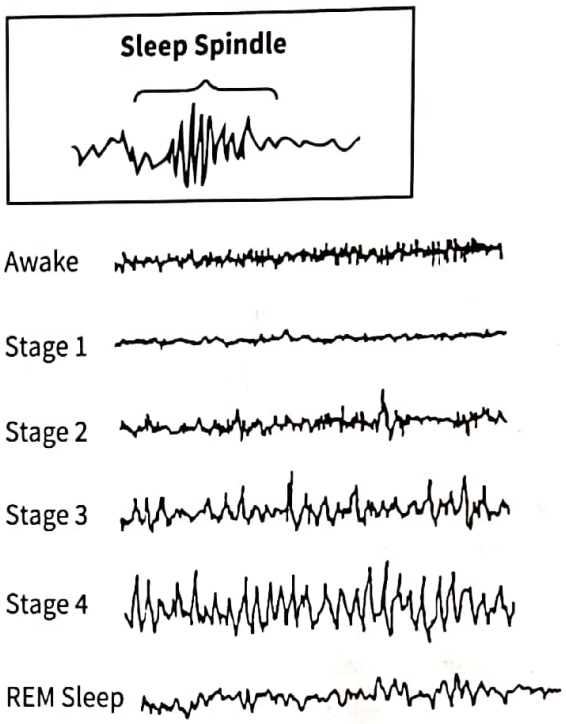
Sleep and dreaming are difficult to study as the participant is not responsive. The sleeper can often give a description of a dream when they awake, but to find out about sleep and dream states, physiological measures are needed too.

Aserinsky and Kleitman (1955) were the first to use physiological measures of sleep to explore the relationship between sleep and dreaming. Like Dement and Kleitman, they used an **electroencephalograph (EEG)** to record brain activity and eye movements and showed that we have several stages during the night, alternating between REM and nREM sleep (see Figure 2.6). They found that participants woken from REM sleep were more likely to report a vivid, visual dream than when woken in other stages. This was the case for both normal and schizophrenic participants.



2.6 EEG recording or sleep stages showing four phases of REM sleep with nREM stages (1–4) in between

An EEG produces a chart (an encephalogram) showing how brain waves vary, i.e. how the **frequency** and **amplitude** (height) of electrical activity changes over time (see Figure 2.7). The chart records changes which indicate the sleep stage a person is in. An EEG can also be used to detect activity in the muscles moving the eyes, so can be used to measure eye movements. This is sometimes called an ElectroOculoGram (EOG).



2.7 EEG recordings: the different stages of sleep

KEY TERMS

internal validity: how well an experiment controls for confounding variables. If an experiment has internal validity the researcher is confident that it is only the IV that is affecting the DV and no confounding variables are having an impact on the results.

rapid eye movement sleep (REM): a stage of sleep in which our eyes move rapidly under the lids, which is associated with vivid, visual dreams.

non-rapid eye movement sleep (nREM): the stages of sleep (1 to 4) in which our eyes are still. It is also called quiescent (quiet) sleep. This is not associated with dreaming.

electroencephalograph (EEG): a machine used to detect and record electrical activity in nerve and muscle cells when many are active at the same time. It uses macroelectrodes, which are large electrodes stuck to the skin or scalp (note they are *recording* electrodes – they cannot give the participant an electric shock!).

frequency: the number of events per fixed period of time, e.g. the number of eye movements per minute (approximately 60/minute in REM sleep) or the number of brain waves (cycles) per second or Hertz (Hz), e.g. 13–30 Hz for beta waves.

amplitude: the 'height' of waves, e.g. on an EEG (indicating voltage).

RESEARCH METHODS

Measures of **frequency** and **amplitude** produce quantitative data. Data such as this tends to be highly **reliable**, as the machines used to produce it, such as the **EEG**, do not vary in the way that they measure the variables. In general, would you expect qualitative data to be more or less reliable? Why?

Reflections: You have just woken up from a dream. It felt as though it had lasted for a very long time. How long do you think you were actually dreaming for?

Method

Research method and design

This study was conducted in a laboratory but several methods were used. To answer the first question above, about the difference in dream recall between REM and nREM sleep, an experiment with a repeated measures design was used. The independent variable was whether the participant was woken from REM or nREM sleep. The dependent variable was whether they recalled a dream or not. The test of question 2, about the relationship between dream duration and the length of the REM period was a correlation (although the comparison between estimates of 5 and 15 minutes was another repeated measures design experiment). To find out about question 3, the relationship between eye movement patterns and dream content, self-reports were compared to the direction of eye movements observed.

Sample

Seven male and two female adults were used, five of whom were studied in detail. The remaining four were used to confirm the results of the first five.

Procedure

On each day of the study participants ate normally, excluding caffeine-containing drinks (such as coffee) and alcohol. They arrived at the laboratory just before their normal bedtime. The participant went to sleep in a dark, quiet room with electrodes attached beside the eyes and on the scalp (the EEG), which fed into the experimenter's room. The wires were gathered together into a single cord from the participant's head (like a pony-tail) so they could move easily in bed (see Figure 2.8).

2.8 The participant slept in a bed with wires from the EEG electrodes leading into the room where the experimenter sat

Participants were woken (by a doorbell) at various times during the night, asked to describe their dream if they were having one, then returned to sleep. They were not told about their EEG pattern or whether their eyes were moving. The procedure for the three questions differed:

1 Participants were woken either from REM or nREM sleep but were not told which. The choice of REM or nREM waking was decided in different ways for different participants:

- using a random number table (participants PM and KC)
- in groups of three REM then three nREM (participant DN)
- by telling the participant that they would *only* be woken in REM but actually waking them in REM or nREM randomly (participant WD)
- in no specific order, the experimenter just chose (participant IR).

Immediately after being woken, the participant stated whether they were having a dream or not and then, if appropriate, described the content of the dream into a recorder. When the participant had finished, the experimenter occasionally entered the room to ask further questions about the dream. There was no other communication between the experimenter and the participant.

2 Participants were woken after either 5 or 15 minutes in REM sleep. The participant guessed which duration they had been dreaming for. Longer REM periods were also allowed. The number of words in the dream narrative was counted.

3 The direction of eye movements was detected using EEG electrodes around the eyes. Participants were woken after a single eye-movement pattern had lasted for more than one minute and asked to report their dream. The eye-movement patterns detected were: 'mainly vertical', 'mainly horizontal', 'both vertical and horizontal' and 'very little or no movement'. Comparison EEG records were taken from awake participants, 20 naive ones and five of the experimental sample, who were asked to watch distant and close-up activity.

The environment was highly controlled, for example the doorbell used to wake participants was sufficiently loud to rouse them immediately from any sleep stage. If the experimenter asked any questions, this was not done until the participant had definitely completed his recording. Also, reports were not counted as 'dreams' if the participant could only recall having dreamt, rather than the content, or had only a vague, fragmented impression of the dream.

Results

Dement and Kleitman reported some general findings, such as that all participants dreamt every night, as well as those relating to their three questions. They found that uninterrupted dream stages:

- lasted 3–50 minutes (with a mean of approximately 20 minutes)
- were typically longer later in the night
- showed intermittent bursts of around 2–100 rapid eye movements.

In addition, they observed that:

- no rapid eye movements were seen during the onset of sleep even though the EEG passed through a stage of brain waves similar to those produced during REM sleep.
- the cycle length (from one REM stage to the next) varied between participants but was consistent within individuals, e.g. between 70 minutes at the shortest and 104 minutes at the longest (with a mean of 92 minutes for all participants).
- When woken from nREM sleep participants returned to nREM but when woken from REM sleep they typically did not dream again until the *next* REM phase (except sometimes in the final REM phase of the night). As a consequence, the pattern of REM and nREM periods was very similar in experimental participants whose sleep was disturbed to those who had an uninterrupted night's sleep.

RESEARCH METHODS

The **mean** sleep cycle length was calculated for each individual. This would have been worked out by adding together the cycle lengths in minutes for every complete cycle a participant had slept through. This would then have been divided by the number of complete sleep cycles that had been observed for that participant. How variable was the average cycle length?

Reflections: Think about occasions when you have woken up in the middle of the night. Are you more or less likely to remember a dream then than when you wake up in the morning? Why?

In relation to their three questions, Dement and Kleitman's results were as follows.

Does dream recall differ between eye movement (REM) and quiescent (nREM) stages of sleep?

Participants frequently described dreams when woken from REM but rarely did from nREM sleep although there were some individual differences (see Table 2.1). Of awakenings from REM, 79.6% (152/191) of awakenings produced dream recall and 93% (149/160) of awakenings from nREM did not produce dream recall.

This difference was most noticeable at the end of the nREM period. In 17 nREM awakenings soon after the end of a REM stage (within 8 minutes), five dreams were recalled (29% of occasions). However, from 132 awakenings following periods longer than eight minutes after a REM stage, only six dreams were recalled (i.e. less than 5% of occasions). In nREM awakenings, participants tended to describe feelings but not specific dream content. They were least likely to remember a dream if they were woken at the stage of sleep

Participant	Rapid eye movements		No rapid eye movements	
	Dream recall	No recall	Dream recall	No recall
DN	17	9	3	21
IR	26	8	2	29
KC	36	4	3	31
WD	37	5	1	34
PM	24	6	2	23
KK	4	1	0	5
SM	2	2	0	2
DM	2	1	0	1
MG	4	3	0	3
Totals	**152**	**39**	**11**	**149**

Table 2.1 Dream recall following awakenings from REM or nREM sleep

in which the EEG has 'spindles' (i.e. stage 2, see Figure 2.7). They tended to be bewildered and report feelings as such as anxiety, pleasantness and detachment.

It is important to note that participant DN was no more accurate even though he could have learned the pattern of awakenings and WD was no less accurate even though he was misled to expect to be dreaming every time. Also, participants did not become any more accurate over time, i.e. they did not improve with practice.

Awakenings from REM sleep did not always produce dream recall, absence of dreaming in REM was more common early in the night. Of 39 REM awakenings when dreams were not reported, 19 occurred in the first 2 hours of sleep, 11 from the second 2 hours, 5 from the third 2 hours and 4 from the last 2 hours. In contrast, awakenings from nREM always produced a low incidence of dream recall.

Is there a positive correlation between subjective estimates of dream duration and the length of the REM period before waking?

The accuracy of estimation of 5 or 15 minutes of REM was very high (88% and 78% respectively). Table 2.2 shows the results of this experimental comparison. REM duration and the number of words in the narrative were significantly positively correlated. The r values for each participant varied from $r = 0.4$ to $r = 0.71$, indicating moderate to strong positive correlations. These were based on between 1 and 35 dreams per participant, with a total of 126 dreams.

One participant often underestimated the dream duration, perhaps because he could only remember the end of the dream. Similarly, narratives from REM periods of recalled after 30–50 minutes of REM were not much longer than those after 15 minutes even though the participants felt they had been dreaming for a long time. This is probably because they could not remember all the details from very long dreams.

Participant	Whether the participants' judgement of 5 or 15 minutes was correct or not			
	5 minutes		15 minutes	
DN	8	2	5	5
IR	11	1	7	3
KC	7	0	12	1
WD	13	1	15	1
MP	6	2	8	3
Total	**45**	**6**	**47**	**13**

Table 2.2 Results of the comparison of dream duration estimates after 5 or 15 minutes of REM sleep

Are eye movement patterns related to dream content?

Eye movement patterns were found to be related to dream content. This part of the study was based on 35 awakenings from nine participants. Periods of only vertical or only horizontal movements were very rare. There were three dreams with mainly vertical eye movements. In one the dreamer was standing at the bottom of a tall cliff operating a hoist (a lifting machine) and looking up at climbers at various levels then down at the machine. In another a man was climbing up a series of ladders looking up and down as he climbed. In the third, the dreamer was throwing basketballs at a net, shooting, looking up at the net, and then looking down to pick up another ball from the floor (Figure 2.9). There was one instance of horizontal movement, in which the dreamer was watching two people throwing tomatoes at each other.

2.9 Dreams of climbing ladders and throwing basketballs produced vertical eye movements

Reflections: Keep a dream diary for a week, writing down the content of any dreams you can remember when you wake up. Think about whether each dream would have had few or many eye movements and whether these would have been vertical or horizontal.

Ten dreams had little or no eye movement and the dreamer reported watching something in the distance or staring at an object. Two of these awakenings also had several large eye movements to the left just a second or two before

the awakening. In one, the participant had been driving a car and staring at the road ahead. He approached a road junction and was startled by a speeding car suddenly appearing to his left (as the bell rang). The other dreamer also reported driving a car and staring at the road ahead. Immediately before being woken he saw a man standing on the left of the road and acknowledged him as he drove by.

Twenty-one of the awakenings had mixed eye movements. These participants reported looking at objects or people close to them, for example talking to a group of people, looking for something and fighting with someone. There was no recall of distant or vertical activity.

The eye movement patterns recorded from the awake (control) participants were similar in amplitude and pattern to those occurring in dreams. Similarly, there were virtually no eye movements when watching distant activity and much more when watching close-up activity. Vertical eye movements were rare in awake participants, except during blinking, and when the experimenter threw a ball in the air.

Conclusions

Dreaming is reported from REM but not nREM sleep, participants can judge the length of their dream duration and REM patterns relate to dream content. As a consequence, dreaming is more likely at the end of the night, as the REM stages are longer. These two observations fit with those reported by other researchers. The occasional recall of dreams from nREM is likely to happen because dreams are being recalled from the previous REM phase (as this is more likely closely following REM sleep). The finding that REM sleep occurs in phases during the night helps to explain why participants in other studies who were awoken randomly may not have reported dreaming. Perhaps they were only woken in nREM stages, or were dreaming about distant objects so had few REMs, making accurate detection difficult. Measurement of eye movements and brain waves has shown that dreams progress in 'real time' and that this is a more objective way to study dreaming than using subjective recall of dreams alone, which can also be affected by forgetting.

Strengths and weaknesses

One method used was the **laboratory experiment**. In this part of the study it was therefore possible to control **extraneous variables**. If some participants, or participants in different stages of sleep, had woken more slowly they may have forgotten more of their dream. This was avoided by using a loud doorbell that woke them instantly, from any sleep stage. The participants were not told about their EEG pattern or whether their eyes were moving in order to avoid possible **demand characteristics**, for example if they expected to remember more detailed dreams in REM sleep they may have made greater effort to do so.

Another method used was a correlation. This demonstrated a positive correlation between REM duration and the number of words in the dream narrative. However, as with any correlation, this could only demonstrate whether there was a link between variables. In order to explore this link in a more controlled way, Dement and Kleitman conducted the additional experiment comparing 5 and 15 minute REM sleep periods.

The definition of a 'dream' was clearly operationalised, as a recollection that included content, rather than just having the impression that they had been dreaming. This helped to raise **validity**, as Dement and Kleitman could be more sure that the details being recorded were of dreams. At the beginning of the study, participants were asked to estimate how long they had been dreaming and although they were roughly (and occasionally exactly) accurate, this task was too difficult, so the method was changed. The task was limited to a choice between 5 and 15 minutes. This also helped to raise **validity** as it reduced **participant variables** such as differences in the ability to recall dreams.

An EEG is an **objective** way to investigate dreaming as it is a biological measure. Differences in narrative length, however, depended not only on the length of the REM phase but also how expressive the participant was, making these reports more **subjective**. Nevertheless, this means that the study collected both **quantitative data**, from the brain waves, eye movement patterns and REM sleep duration, and **qualitative data**, which helped to provide insight into the reason for the eye movements detected. The EEG also provides a very reliable measure because it is unaffected by the experimenter's personal view. The consistent placing of the electrodes ensured that recordings taken from each participant would provide the same information. The reliability of the findings is supported by the similarity of the results to those of previous studies.

As there may be differences in the dreaming of men and women, or between the way they report their dreams, it was useful in terms of **generalisability** that there were both genders in the sample. However, as there were only nine participants in total, the small size of the **sample** limits generalisability.

One aspect of the method that raised an **ethical issue** was the **deception** of participant WD who was misled about the stage of sleep he was being woken in. Participants should not be deceived as it can cause distress and means they cannot give their **informed consent**. However, in some

cases the aim cannot be achieved without doing so and in this case it provided a way to test whether expectation of being woken in REM (at least sometimes) would affect a participant's dream reports.

Several aspects of the procedure potentially reduced the ecological validity of the findings. People who were used to drinking coffee or alcohol could have experienced sleep or dreams that were not typical for them as they had been asked to refrain from those drinks. Also, all participants would have found sleeping in a laboratory, connected to machines and under observation, quite different from sleeping in their normal bed. This could also have made their sleeping behaviour less typical.

SELF-ASSESSMENT QUESTIONS

4 Biyu is planning a study on dreams and is worried that if her participants know the aim, they might make dreams up to please her.
   a Explain why this would be a problem.
   b She had decided to solve this problem by telling her participants the study was about insomnia, but her teacher says she cannot do this. Explain why Biyu's teacher has said this.

5 Karl is aiming to find out whether people sleep for longer because they have eaten more. He plans to do a correlation, asking people how much they have eaten during the day and how long they sleep for that night. Karl's teacher says this will not work.
   a Suggest why Karl will not be able to use the information he obtains from his study to come to a conclusion about his aim.
   b Karl decides to do the study anyway. Explain how he could operationalise the variables of sleeping and eating.

Summary

In Dement and Kleitman's study an EEG was used to collect data about sleep stages and eye movements and details of dream content were obtained by ***self-report****. This study was conducted in a* ***laboratory experiment****, allowing for controls such as over what participants were told and how they were woken and objective, quantitative records could be collected as well as* ***qualitative data****. Three questions were answered: dreams are reported from REM but not nREM sleep, we can accurately judge the length of dreams and REM patterns relate to dream content.*

2.3 Core study 3: Schachter and Singer (1962)

Schachter, S., and Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69*, 379–399.

Aim

The aim of the study was to test the Two-Factor Theory of Emotion. Schachter and Singer (1962) wanted to research if, given a state of physiological arousal for which the individual has no adequate explanation, cognitive factors can lead the individual to describe their feelings with any of a number of emotional labels.

Background

How individuals use either internal or external cues to identify their own emotional state has been of interest to psychologists since the late 1800s. Some of the early research suggested that cognitive factors could influence our emotional state. Cognition can be defined as the mental processes of acquiring and processing knowledge and understanding through experiences, senses and thought. Emotion can be defined as the body's adaptive response to a particular situation.

Following on from this it was suggested that an emotional state may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal. The cognition therefore steers our interpretation of our physiological state so that we can label our emotional responses.

For example, if a criminal tries to mug you by pointing a gun and demanding you give him your wallet, you will experience physiological arousal of the sympathetic **nervous system** (Figure 2.10). Your heart rate and respiration rate will increase, your pupils will dilate and **adrenalin** will be released (this is sometimes referred to as

KEY TERMS

nervous system: the brain, spinal cord and all the nerve cells in the body that communicate to control our thinking, behaviour and emotions.

adrenalin: a hormone released from the adrenal glands in response to stress or excitement. It is also known as epinephrine, and is a medication, hormone and neurotransmitter. Common side effects include shakiness, anxiety, sweating; a fast heart rate and high blood pressure may occur. Strong emotions such as fear or anger can cause epinephrine to be released into the bloodstream.

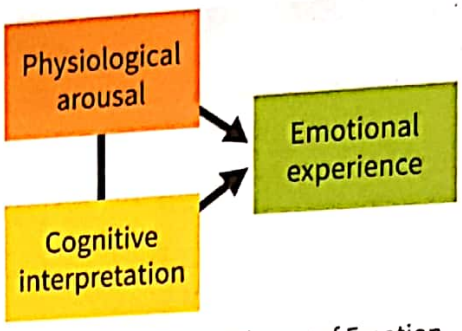
the 'fight or flight' response). You will label this emotional experience as 'fear' because of your knowledge (cognitive interpretation) about criminals with guns and how dangerous they are. Without the knowledge of the danger of guns and criminal behaviour your emotional experience would be different.

What if you were just sitting at home and experienced the same physiological arousal, meaning that you didn't have an explanation for your feelings? Schachter and Singer (1962) suggested that such an individual would look to the situation they were in to gain an explanation.

2.10 Emotional response is a combination of physiological arousal and cognitive understanding of the situation

Schachter and Singer developed what they named the Two-Factor Theory of Emotion (Figure 2.11). They suggest that emotional experience comes from a combination of a physical state of arousal and a cognition that makes best sense of the situation the person is in. For example, the Two-Factor Theory of Emotion argues that when people become aroused they look for cues as to why they feel the way they do from the environment and interpret their arousal in relation to this. Therefore, any emotional experience is a combination of physiological arousal and a cognitive interpretation.

The study by Schachter and Singer investigated emotions, and their Two-Factor Theory argues that we experience emotions as an interaction between physiological and psychological variables. To achieve this, they created a situation where some participants were physiologically aroused through an **epinephrine** injection and put into either a 'happy' or 'angry' situation to investigate if those participants would look to the situation for context cues to explain their arousal.



2.11 The Two-Factor Theory of Emotion

Method

Research method and design

This was a laboratory experiment as the environment in which the participants were tested was not comparable to an everyday situation. It was a highly standardised procedure and all participants were exposed to the same environment with a scripted response from the stooge. There were two independent variables in this experiment which resulted in seven different conditions. The first independent variable concerned the knowledge about the injections; whether they were informed, misinformed or ignorant. The second independent variable was the emotional situation that the participant was placed into following the injection; either euphoria or anger. There was also a control group who were not injected with epinephrine but a saline solution. This experiment is an example of an independent groups design as participants only took part in one of the seven groups (see Table 2.3).

There were two key measures of the **dependent variable**. The first was observational data that was recorded by two observers through a one-way mirror during the emotional arousal element of the experiment. The observer had to measure to what extent the participant acted in a euphoric or angry way. During each stage of the stooge's routine the observer kept a record of how the participant reacted and what was said. In the euphoric condition, the categories into which the behaviour was coded were that the participant: joins in the activity, initiates a new activity, watches stooge or ignores stooge. In the anger condition, the observers coded behaviour into six categories depending on the participants' response. These categories were that the participant: agreed with a comment, disagreed with a comment, was neutral to a comment, initiates agreement or disagreement, watches or ignores the stooge. To test the **reliability** of these measures two observers coded their observations independently. The observers completely agreed on the coding of 88% of the behaviour they saw.

The second was a self-report that the participants completed following the emotional arousal element of the research.

	Euphoria Informed	Anger Informed
Experimental groups	Misinformed	–
	Ignorant	Ignorant
Control group	Placebo	Placebo

Table 2.3 Conditions within the experiment

RESEARCH METHODS

A **dependent variable** is the factor which is measured in an experiment and which the experimenter expects to be influenced by the independent variable. In this case, there two dependent variables, and they are measured in several different ways. Write out the two dependent variables and list the ways they are measured for each one.

Sample

The 185 participants were all male college students taking classes in introductory psychology at the University of Minnesota, of which many received course credit for taking part in the study. The university health records of all participants were checked prior to the experiment to ensure that no harmful effects would result from the injections.

Reflections: Why might psychologists use psychology students in their experiments? What impact might this have on the validity of any findings?

Procedure

When the participants arrived at the laboratory they were told that the aim of the experiment was to test the effects of vitamin supplements on vision:

> *'In this experiment we would like to make various tests of your vision. We are particularly interested in how certain vitamin compounds and vitamin supplements affect the visual skills. In particular, we want to find out how the vitamin compound called "Suproxin" affects your vision. What we would like to do, then, if we can get your permission, is to give you a small injection of Suproxin. The injection itself is mild and harmless; however, since some people do object to being injected we don't want to talk you into anything. Would you mind receiving a Suproxin injection?'*

RESEARCH METHODS

The participants were misinformed about the aims and procedures of the study, ie they were **deceived**. This means they could not give their informed **consent**. Note that these two uses of 'misinformed' and 'informed' are different from the way these words are used in describing the conditions in this experiment. Can you define the ethical issue of 'deception' and the ethical guideline of 'consent'?

A doctor then entered the room, repeated these instructions took the participant's pulse and injected the participant with 'Suproxin' (Figure 2.12). The participants were actually injected with either adrenalin or a **placebo** (saline solution). The dosage that the participants were injected with should have caused them to experience the side effects within three to five minutes and these could last up to an hour.

2.12 Epinephrine or a saline solution was injected into participants. Could this have caused arousal regardless of the content?

Those participants who were injected with adrenalin were then put into one of three experimental conditions: informed, ignorant or misinformed. The fourth condition consisted of those participants who had received the placebo injection.

Participants in the *informed* condition were told that they could expect some side effects of the injection and these were that 'your hand will start to shake, your heart will start to pound, and your face may get warm and flushed'. These are the side effects of an injection of adrenalin so the participants would have an explanation for any feelings.

KEY TERMS

reliability: the extent to which a procedure, task or measure is consistent, for example that it would produce the same results we the same people on each occasion.

placebo: a pill or injection given which the patient or participant believes is a drug but which, in reality, has no effect.

In the *misinformed* condition the participants were told that they could expect some side effects of the injection and these were that 'your feet will feel numb, you will have an itching sensation over parts of your body, and you may get a slight headache'. These are not the side effects of an injection of adrenalin so the participants would not have any explanation for the actual side effects they would experience. This condition was introduced as a control condition.

Those participants in the *ignorant* condition were told that they would experience no side effects from the injection. Therefore, these participants would also not have an explanation for the actual side effects that they would experience as a result of the injection of adrenalin.

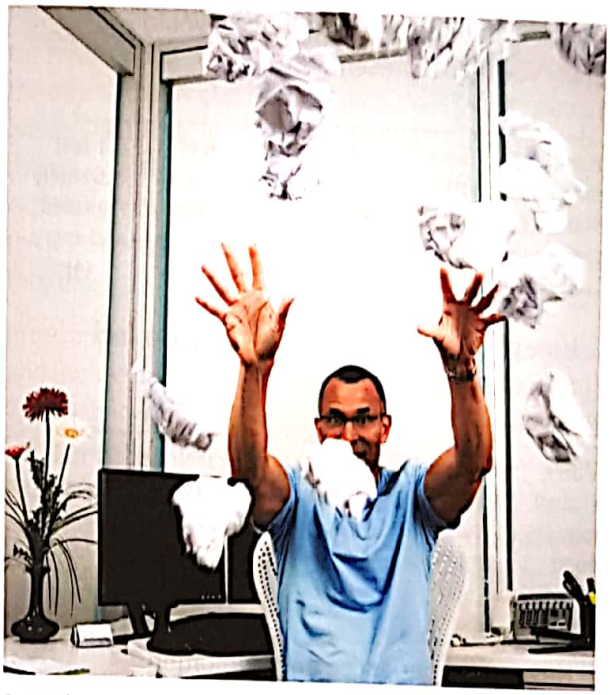
Immediately after the participant's injection, the doctor left the room and the experimenter returned with a **stooge**. The stooge was introduced as another participant and the experimenter stated that both had had the Suproxin injection which would take 10 minutes to be absorbed into the bloodstream, after which they would both be given the same tests of vision. Participants were then exposed to one of two emotional states at this point: euphoria or anger.

In the euphoria condition when the experimenter departed, he apologetically added that if they needed any rough paper, rubber bands or pencils they should help themselves. The waiting room had been arranged to look in a state of disarray. As soon as the experimenter left the room the stooge introduced himself again, made a few ice-breaker comments and then began his routine which consisted of playing with items (paper, rubber bands, pencils, folders and hula hoops) left in the room (Figure 2.13). The stooge suggested that the participant join in while he used the items. The routine was standardised as far as was possible. The stooge never knew which condition any particular participant was in.

KEY TERMS

stooge: a person who appears to be another participant or someone not related to the study, but who is in fact working on behalf of the researcher. They are also sometimes known as '*confederates*' and may be used to mislead real participants within the study.

demand characteristics: features of the experimental situation which give away the aims. They can cause participants to try to change their behaviour, e.g. to fit with their beliefs about what is supposed to happen, which reduces the validity of the study.



2.13 The stooge in the euphoria condition threw paper to create a fun atmosphere

In the anger condition, after the injection the experimenter introduced a stooge and explained that it was necessary to wait 20 minutes to let the Suproxin enter the bloodstream and that the participants had to complete a questionnaire (Table 2.4) during this time. The stooge was instructed to create a feeling of anger in the room and this was achieved through a variety of comments that he made as the 20 minutes passed:

'...it's unfair for them to give you shots'
'This really irritates me'
'The hell with it!'
'I'm not wasting any more time'

These increased in intensity and were linked with the questions in the questionnaire. As the participants worked through the questionnaire the questions became more personal, and the stooge became increasingly irate in his behaviour.

7	List the foods you would eat in a typical day.
13	List the childhood diseases you have had and the age at which you had them.
28	How many times a week do you have sexual intercourse?
34	With how many men (other than your father) has your mother had extramarital relationships?

Table 2.4 Example questions asked in the 'anger' condition

How irritated, angry or annoyed would you say you feel at present?

I don't feel at all irritated or angry (0)	I feel a little irritated and angry (1)	I feel quite irritated and angry (2)	I feel very irritated and angry (3)	I feel extremely irritated and angry (4)

How good or happy would you say you feel at present?

I don't feel at all happy or good (0)	I feel a little happy and good (1)	I feel quite happy and good (2)	I feel very happy and good (3)	I feel extremely happy and good (4)

Have you experienced any palpitation (consciousness of your own heart beat)?

Not at all (0)	A slight amount (1)	A moderate amount (2)	An intense amount (3)

Did you feel any tremor (involuntary shaking of the hands, arms or legs)?

Not at all (0)	A slight amount (1)	A moderate amount (2)	An intense amount (3)

2.14 Self-report questions that the participants completed at the end of the experiment

When the participant's session with the stooge was complete, the experimenter returned to the room, took their pulses, and told them there was one final questionnaire that considered their physical responses to the Subterfuge (Figure 2.14). This was used as the self-report measure for the dependent variable. When the participants had completed these questionnaires, the experimenter announced that the experiment was over, explained the deception and its necessity in detail, answered any questions and swore the participants to secrecy to protect future runs of the experiment. All participants gave consent to take part in the study and the researchers checked their medical records to ensure that the injections that were given would not cause any harm. To further protect the participant, the injection was administered by a trained doctor who was around for the duration of the experiment to monitor the participants. The researchers did deceive the participants in this research as they were not aware of the true content of the injection, but this was necessary to prevent **demand characteristics** and improve the validity of the results.

Hypotheses

1 If a person experiences a state of arousal for which they have no immediate explanation, they will label this state and describe their feelings in terms of the cognitions available to them at the time.

2 If a person experiences a state of arousal for which they have an appropriate explanation (e.g. 'I feel this way because I have just received an injection of adrenalin'), then they will be unlikely to label their feelings in terms of the alternative cognitions available.

			Physical measure	Self-report measures	
		N	Change in pulse	Palpitations	Tremors
	Euphoria				
Adrenalin (Epinephrine)	Informed	25	2.9	1.2	1.4
	Ignorant	25	1	1.8	1.8
	Misinformed	25	3.1	1.3	2.0
Control	Placebo	26	−3.3	0.3	0.2
	Anger				
Adrenalin (Epinephrine)	Informed	22	6.5	1.3	1.4
	Ignorant	23	11.8	1.4	1.8
Control	Placebo	23	−4.9	0.6	0.2

Table 2.5 Average scores for the seven conditions

3 If a person is put in a situation, which in the past could have made them feel an emotion, they will react emotionally or experience emotions only if they are in a state of physiological arousal.

Results

All but one of the 185 participants agreed to the injection. Eleven participants were so suspicious about some crucial feature of the experiment that their data was discarded. Five participants had no physiological reaction to the injection of epinephrine and were also excluded from the data analysis. This left 169 participants' data to be analysed (Table 2.5).

The participants who received the injections of adrenalin showed significantly more **sympathetic arousal** (as measured by pulse rate and self-ratings on numbness, tremor, itching, palpitation and headache) compared with the placebo participants. The misinformed condition was only run in the euphoria condition as it was a control condition and inclusion in just one of the categories was adequate to evaluate the possible impacts of receiving information about side effects after the injection. From the data in Table 2.5 it can be immediately seen that on all items those participants in the adrenaline condition show more evidence of physiological response (change in pulse rate) compared to the control groups. Further to this, on the self-report measures, those participants who were in the adrenaline conditions reported higher scores for palpitations and tremors suggesting that they were having a behavioural response to the increased levels of arousal.

In all the adrenalin conditions pulse rate increased significantly when compared with the decreased characteristic of the placebo conditions. On the self-report scales about palpitations and tremors it is clear that the participants in these conditions experienced more of these symptoms in comparison to the placebo conditions. The difference between the adrenalin conditions and the placebo conditions are all significant at the 0.001 level of significance. Therefore, it is clear that those participants in the adrenalin conditions were physiologically aroused during the experiment.

KEY TERM

sympathetic arousal: when we are exposed to a stressful situation, the sympathetic nervous system becomes aroused causing the pupils to dilate, an increase in heart rate, digestive activity is inhibited and glucose is released by the liver for extra energy needed to prepare the body to respond to alarm or stress.

From the self-report measures, in the euphoria condition, the misinformed participants were feeling happier than all the others; the second happiest group was the ignorant group. This demonstrates that these participants were more susceptible to the stooge because they had no way of explaining why they felt as they did. The informed group felt the least positive because they were aware why they felt as they did.

In the anger condition, the ignorant group felt the angriest and the second angriest group was the placebo group. The least angry group were those who had been informed. This shows that participants were more susceptible to the stooge because they had no way of explaining why their body felt as it did. Behaviour of the groups was observed through a one-way mirror and matched their self-reports.

Conclusions

Schachter and Singer argue that their findings support their Two-Factor Theory of Emotion, which states that the physiological arousal in different emotions is entirely the same and we label our arousal according to the cognitions we have available. They argued that all three of their hypotheses were supported and concluded that if a person experiences a state of arousal for which they have no immediate explanation, they will label this state and describe their feelings in terms of the cognitions available to them at the time. This study helps us understand how people use different environmental cues to help them interpret their physiological state. It could be useful in treating people who suffer with anxiety and panic attacks as it will allow them to identify environmental triggers that may cause them to become physiologically aroused and fearful.

The Two-Factor Theory of Emotion has been an influential theory of emotions; however subsequent work has shown that the relationship is more complex than this theory predicts. More recent work suggests that attempts to understand an unexplained state of arousal is more extensive than a quick examination of cues in the surrounding environment. When an individual seeks to explain their state of arousal, they do not just use the behaviour of those around, but also call on many other sources of information such as past experience and knowledge about the context of any behaviour.

Reflections: How often do you look to the situation to explain how you feel?

How might information you know influence the way that you feel?

Schachter and Singer's laboratory experiment exerted a large amount of control through their standardised procedure. For example, they were able to randomly allocate different participants to the different conditions, they were able to deceive the participants of the real nature of the experiment and standardise the procedure as much as possible. They even ensured that the stooge did not know which condition the participant was in (**double blind technique**).

KEY TERM

double blind technique: when both the participant and the researcher are unaware of which condition the participant is in to prevent demand characteristics and act as a control to improve the validity of any data collected.

The sample in this research consisted of university students and introduced participant variables that could distort the outcomes of the research reducing the validity of the research. Adrenalin does not affect everyone in the same way, for example five of Schachter and Singer's participants were excluded from analysis because they experienced no physiological symptoms. A further problem with the procedure is that no assessment was made of participants' mood before the injection to gain a baseline from which to measure any change as a result of the stooge. Also, the sample consists of only male participants which further impacts on our ability to generalise the results to all; males could experience emotion differently to females and this has been ignored in this research. People do not usually experience emotions in the way in which Schachter and Singer induced them and are often aware of events before the onset of arousal giving us information to interpret our physical cues.

The use of **questionnaires** as a method to **operationalise** the **dependent variable** allowed Schachter and Singer to further standardise the collection of the data within the experiment. Questionnaires allow researchers to collect **quantitative data** quickly from participants that can be easily analysed and used to compare results across two conditions. Completing a questionnaire may not give the participants the same freedom to express their feelings as a discussion therefore we could argue that there is little ecological validity in this research. However, Schachter and Singer also use other measures such as observation and physical measures to assess both the physical and psychological impact of the arousal.

SELF-ASSESSMENT QUESTIONS

6 Explain why the results support the Two-Factor Theory of Emotion.

7 Explain why the experiment could be seen as unethical.

8 Describe **two** ways in which the researchers measured the emotional response of the participants.

9 Evaluate the measures used by Schachter and Singer to measure the emotional response of the participants.

2.4 Issues, debates and approaches

The application of psychology to everyday life

The use of the EEG allows psychologists to accurately detect dreaming. This could have useful applications for people with sleep and dream disorders, for example people with insomnia or nightmares or people who sleep walk.

Nature versus nurture

The biological approach focuses mainly on the nature side of this debate, which is why it is possible to obtain evidence through procedures like the EEG, which was used in the Dement and Kleitman study to measure brain waves and eye movements. It is useful to be able to collect physiological evidence about brain activity as it provides direct evidence for the underlying biological processes, such as the link between dream content and eye movements.

Dream content relates to our experiences, so is a product of nurture. This, at least partly, explains the differences in dreams between individuals. Nurture influences will vary, thus the content of people's dreams will differ. However, as even a foetus in the uterus experiences REM sleep, the capacity to dream appears to be a product of nature.

Similarly, the biological processes underlying emotions are the product of the brain, and of **hormones**. However, there are clear differences between us in terms of our emotional responses. These can be accounted for by both differences in nature, such as hormone levels, and in terms of our experiences, i.e. nurture.

KEY TERM

hormones: chemicals that are released from glands and travel around the body in the blood to communicate messages between organs.

Using children and animals in psychological research

Although none of the studies in the section used children or animals, it is useful to consider whether they could have been used and what the findings might have shown. For example, there is some debate about whether animals show emotions at all. They certainly show responses such as fear, for example in the presence of a competitor or predator. However, studies exploring such responses are likely to be unethical. The expression of 'feelings' in animals have been explored in more ethical ways, for example by seeing the choices that hens make when presented with situations they do and do not like (Bubier, 1996, quoted in Stamp-Dawkins, 1998).

Reflections: It is possible to use EEGs on the brains and eyes of babies. Could you devise a study to test whether dream content is related to eye movements in babies? If so how, if not, why not?

Non-human animals such as mammals do demonstrate REM sleep, and this has been studied extensively. It is very difficult, however, to determine the content of their dreams. This has been done indirectly, by using brain-cell recording techniques to explore repeated waking behaviours (such as bird song and rats running mazes). It has been found that the patterns of brain activity during sleep closely resemble those of the waking behaviours, indicating what the animals are dreaming about.

Individual and situational explanations

Our tendency to cry at sad films clearly suggests that situational factors matter in our expression of emotions. Although individual factors are present here too – not everyone cries at the same film. Differences such as these were found in both the study by Canli et al. and in Schachter and Singer, for example in the extent to which participants were affected by the mood of the stooge.

Summary

Canli et al.'s study investigated the brain response to emotive images and how they affect memory. Participants had fMRI scans which showed more amygdala activation for emotive than non-emotive images. These images also felt more emotionally arousing to the participants and were better recalled. This laboratory study was well controlled with standardised exposure time and emotional stimuli. Although an fMRI lacks ecological validity and may, in itself, affect emotional responses, participants' scans cannot be affected by demand characteristics. The use of a repeated measures design also ensured that individual differences in emotional responses, brain activity or recall ability could not affect the results. However, the all-right handed, female sample means that the results may not generalise to males or left-handers if they differ in emotional responses.

Dement and Kleitman's study explored the relationship between eye movements during sleep and dream recall. An EEG provided information about participants' sleep stages, such as REM sleep, and about eye movements. Dream recall was measured by self report. The results showed that dreams occurred in REM rather than nREM sleep, that the direction and amount of eye movements during dream sleep is related to dream content, e.g. vertical movements of the eyes and dream events, and that estimates of dream duration are generally accurate. Many variables, such as food and drink that could affect sleep, were controlled and demand characteristics were reduced where possible. However, the laboratory and equipment meant that the situation was unusual for sleeping. Nevertheless, objective, quantitative data was obtained from the EEG, eye movements and timing, allowing for valid and reliable comparisons to be made.

Shachter and Singer tested the two-factor theory of emotion by manipulating physiology with injections of adrenalin and cognition with information (informed, misinformed or not informed) and using an angry or euphoric stooge. The manipulation of cognitions gave the participants different explanations for the physiological effects they were feeling. The results showed that adrenalin did increase physiological arousal (e.g. raising pulse rate) but that this was only interpreted as anger/euphoria when participants had no existing reason for how they felt (if they were misinformed or not informed). This is because they needed a cognitive explanation so took it from the situation (the stooge), so the results support the two-factor theory. The results may not be generalisable, because of individual differences (some participants were not affected by the injection) and because no females were tested. The participants may also have differed prior to the experiment, and this was not measured. Nevertheless, the use of both misinformed and not informed controls, and two different emotional conditions (anger and euphoria), adds validity to the study, as does the range of measures used.

Exam-style questions

1 Describe the variables that were measured in the study by Canli et al. **[6 marks]**

2 Describe **two** conclusions from the study by Canli et al. **[2 marks]**

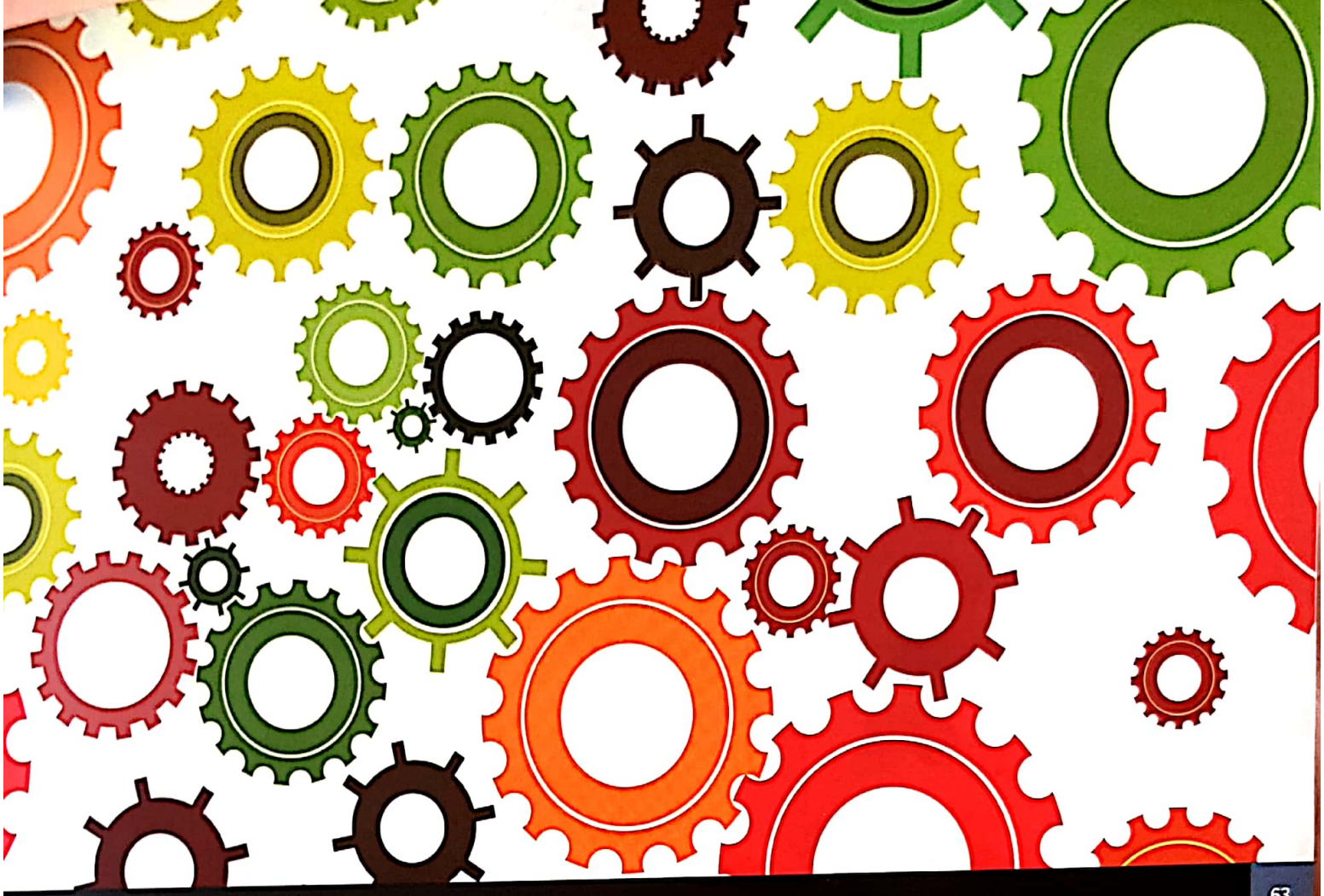
3 One strength of the biological approach is that it is scientific. Describe **one** way in which the study by Dement and Kleitman is scientific. **[2 marks]**

4 **a** Describe **one** aim from the study by Dement and Kleitman. **[1 mark]**

b Explain how the results support this aim. **[3 marks]**

5 Explain **two** ways in which the biological approach is different from the learning approach. Use the study by Schachter and Singer as an example of the biological approach. **[4 marks]**

6 There are many ethical problems with the study by Schachter and Singer but also some ethical strengths. Describe **one** ethical **strength** of this study. **[2 marks]**



Chapter 3
Cognitive approach

Introduction

The aim of this chapter is to introduce you to the cognitive approach in psychology and to explore three studies from this approach. They are:

- **Andrade (doodling)** which suggests that doodling can improve concentration and the memory of a conversation
- **Baron-Cohen et al. (eyes test)** which investigates how a lack of a 'theory of mind' in adults with Asperger's Syndrome or autism can result in problems recognising emotions
- **Laney et al. (false memory)** which explores how false memories can impact on memories and beliefs in relation to eating asparagus.

These three studies illustrate the main assumptions of the cognitive approach which suggests that:

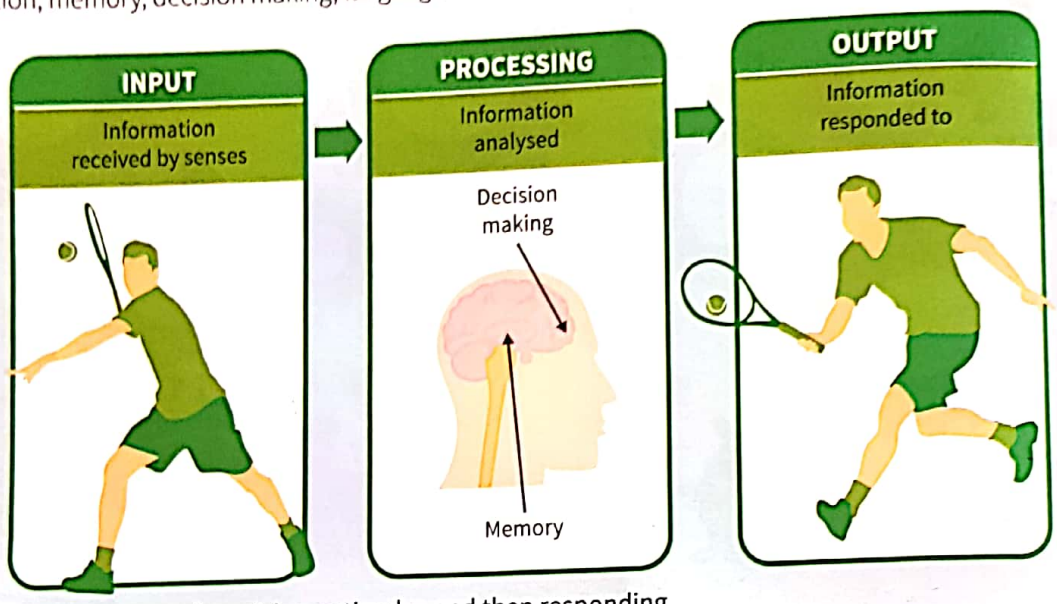
- behaviour and emotions can be explained in terms of cognitive processes such as attention, language, thinking and memory
- similarities and differences between people can be understood in terms of individual patterns of cognition.

This chapter will examine how different cognitive processes can have an impact on a person's behaviour. In each core study you will learn about the background to the research, the way it was conducted, the results and the conclusions. You will also see how different research methods are used within the cognitive approach and how a range of issues and debates can be applied to the concepts considered in this approach.

Is behaviour all about information processing?

Cognitive psychologists are interested in the processes that work within the mind and how these affect our behaviour. Cognitive psychologists study concepts such as attention, memory, decision making, language development and how cognitive abnormalities can impact on how we behave.

The mind actively processes information that comes in from our different senses. Cognitive processes mediate between stimulus and response. Sometimes this process is compared to a computer and discussed in terms of INPUT > PROCESS > OUTPUT.

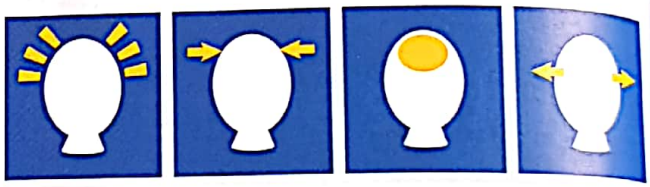


3.1 The process of receiving a stimulus and then responding

For example, when playing a sport our senses process information about other team members, or where the ball is and where we are in relation to it. This is the INPUT element of cognition. All of this sensory information is then processed in the brain with some areas (such as the prefrontal cortex) making decisions about necessary actions, often accessing memories to help inform the process. This is the PROCESS. As a consequence of this process we move, respond and act in an appropriate way. This is the OUTPUT. All of this usually occurs as a seamless subconscious process (Figure 3.1).

Memory is a cognitive process. As a student you spend a lot of time using your memory and hoping that when it really counts you can recall that important piece of information in the exam. Without memory, learning could not take place. Memory is the cognitive ability to encode, store and retrieve information.

When we experience an event, we have to encode (convert) this sensory information into a format that the brain can store. This is the brain's ability to maintain a record of the event without actively using it. Finally, when you need that piece of information, you need to be able to recall, or retrieve, the information (Figure 3.2).



3.2 The memory process

Reflections: How much information have you processed in the last five minutes? What about the last five hours? How much of the information that you have processed have you been aware of and can you recall? You can probably remember a general summary of what you've just read, but could you write it word-for-word?

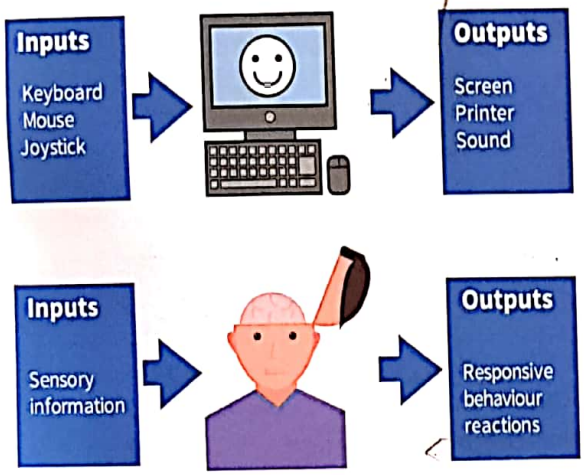
If you think about what you ate six days ago, can you remember it? You should be able to, since you were there and all your senses were involved: you saw your food, you tasted and smelled it and probably talked about it with those around you. Can you remember? Probably not.

Now consider this: if you forget something, is it gone forever? What might you need to help you recall what you ate six days ago?

3.1 The cognitive approach

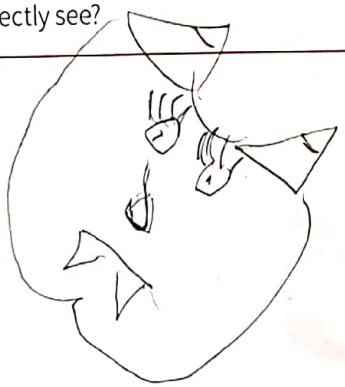
In this chapter we will be looking at how cognitive process like memory can be influenced by doodling and focusing concentration. This suggests that there are ways that we can support our cognitive processes to improve our memories of events. We will look at autism, a problem with cognitive processing that can influence a person's ability to interact with people socially and impair their **social cognition**. Finally we will see how false memories can influence our preferences for different foods.

The cognitive approach often adopts scientific procedures to develop and test hypotheses using experimental techniques. Cognitive theories simplify cognitive processes and allow us to understand mental processes that are not directly observable. However, the cognitive approach tends to ignore biology and genetic influences and provides a mechanistic view of human behaviour (Figure 3.3).



3.3 Cognitive psychologists often compare the way the brain processes information to a computer

Reflections: Can such a scientific approach to understanding mental processes really tell us about how we think, feel and behave? What issues might we encounter trying to measure cognitive processes that we cannot directly see?



KEY TERM

social cognition: the study of how people process social information and how this processing might affect how a person behaves towards or around other people.

3.2 Core study 1: Andrade (doodling)

Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology, 24*(1), 100–106.

Aim

Doodling, such as shading in all the Os in 'psychology' on a worksheet, is a common activity. Andrade was interested to know whether this activity assisted information processing, perhaps by enabling people to attend more effectively or by enhancing their memory.

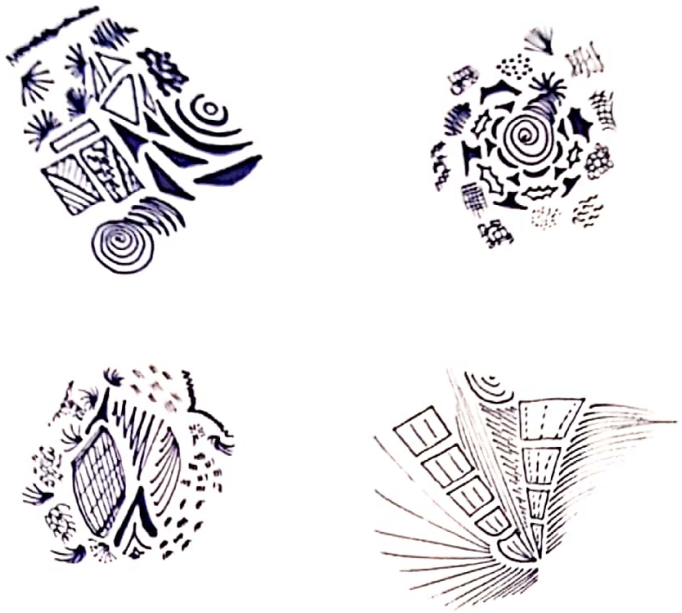
Background

Reflections: The lesson is hard but you're paying **attention**. Your teacher stops talking and asks you what you are writing. You say 'nothing' only to discover a neatly coloured-in shape on the front of your file... You're on the phone and think you're listening properly, but you've doodled an enormous, elaborate leaf pattern during the call. Do you think this doodling is helping you to concentrate or making it more difficult (Figure 3.4)?

KEY TERM

attention: the concentration of mental effort on a particular stimulus. It may be focused or divided.

Teachers often tell students off for doodling. 'I didn't even know I was doing it', you think to yourself in self-defence. Your teacher's complaint is justified: research has shown that we



3.4 Doodling – does it help or hinder concentration?

perform less well when our **attention is divided** between tasks. It might make sense to assume that if we are engaged in doodling, we would not be attending as well to any other task, so be worse at them than if we were not doodling. However, in your defence, doodling might be useful. Doodling might aid concentration (Do and Schallert, 2004), for example by reducing **daydreaming** so that you stay **focused**. This idea is based on the **working memory model**. Daydreaming is linked to high arousal when we are bored and it uses important cognitive processing resources (the 'central executive') so inhibits performance on tasks that use this resource – including attention and memory. In Andrade's study, the primary task of listening to a message, was an auditory task whereas doodling is a visuo-spatial task. It is therefore possible that the concurrent task of doodling would interfere less with overall processing than devoting a greater amount of central executive function to daydreaming.

Alternatively, doodling may help to maintain arousal (Wilson and Korn, 2007), for example by giving you something physical to do while you think. It could raise arousal to help to keep you awake if you are sleepy or reduce arousal if you are agitated because you are bored.

Andrade defines doodling the sketching of patterns and figures that are unrelated to the primary task. Such doodling either could take cognitive resources away from the intended (primary) task as if it placed simultaneous demand on cognition by dividing attention, or it could, as would be the case for the most concurrent cognitive tasks, improve performance by raising arousal and enhancing focused attention on the primary task.

KEY TERMS

divided attention: the ability to split mental effort between two or more simultaneous tasks (called 'dual tasks'), for example, driving a car and talking to a passenger. Divided attention is easier when the tasks involved are simple, well practised and automatic. You may notice that inexperienced drivers find conversing more difficult, and any driver may stop mid-sentence if the road conditions become difficult. The primary task is the main task, in this case driving, and the additional task is called the concurrent task because it is happening at the same time.

daydreaming: a mildly altered state of consciousness in which we experience a sense of being 'lost in our thoughts', typically positive ones, and a detachment from our environment.

focused attention: the picking out of a particular input from a mass of information, such as an array or a continuous stream, for example, concentrating on your teacher's voice even when there is building work outside and the student next to you is whispering.

working memory model: This model of memory suggests that two different types of current or 'working' memory can be used at the same time, one is vis spatial and the other auditory. These are governed by an overall 'central executive'.

Method

Research method and design

This was a laboratory experiment; the environment was not the normal place in which people would respond to telephone messages and the situation was controlled. The design was independent measures as participants were either in the control group or in the doodling group.

Sample

The participants were 40 members of a participant panel at the Medical Research Council unit for cognitive research. The panel was made up of members of the general population aged 18–55 years and they were paid a small sum for participation. There were 20 participants in each group, mainly females, with two males in the control group and three in the doodling group (one participant in this condition did not doodle and was replaced).

Procedure

All participants listened to a dull (mock) telephone call about a party.

> 'Hi! Are you doing anything on Saturday? I'm having a birthday party and was hoping you could come. It's not actually my birthday, it's my sister **Jane's**. She'll be 21. She's coming up from *London* for the weekend and I thought it would be a nice surprise for her. I've also invited her boyfriend **William** and one of her old school friends, **Claire**, but she doesn't know that yet. Claire's husband Nigel was going to join us but he has just found out that he has to go to a meeting in *Penzance* that day and won't be back in time. I thought we could have a barbecue if the weather is nice, although the way it has been so far this week, that doesn't look likely. I can't believe it has got so cold already. And the evenings are really drawing in aren't they? Anyway, there is plenty of space indoors if it rains. Did I tell you that I have redecorated the kitchen? It is mainly yellow—the wallpaper is yellow and so is the woodwork, although I thought it would be better to leave the ceiling white to make it look lighter. I've still got the old blue fittings—they are pretty battered now but I can't afford to replace them at the moment. Do you remember **Craig**? I used to share a flat with him when we were both working for that bank in *Gloucester*. He has bought a house in *Colchester* now but he promises to take time off from gardening to come to Jane's party. **Suzie** is going to be there too. She's the person I met at the pottery class in *Harlow* last year. Apparently she has got really good at it and may even be having an exhibition of her work soon. Will you be able to bring some food? Maybe crisps or peanuts, something along those lines. **Jenny** from next door is going to bring a quiche and I'll do some garlic bread. I found a good recipe for punch—you warm up some red wine with gin and orange juice plus cloves and cardamom and cinnamon. Add some brown sugar if it's not sweet enough. The boys from the house down the road have promised to bring some of their homebrew. There are three of them sharing that house now—John, Tony and Phil. I think they were all at college together. Phil teaches at a primary school in *Ely* now and the other two commute to *Peterborough* each day. I think they both work in the hospital there—I know Tony was training to be a nurse at one point so maybe he is qualified now. John can't come on Saturday because his parents are coming to stay for the weekend but **Phil** and **Tony** should be there. Tony has to pick their cat Ben up from the vet so he may be a bit late. By the way, did I tell you about our holiday in *Edinburgh*? It was a complete disaster. We were camping and it rained constantly. We spent most of the time in museums, trying to keep dry and then, to make matters worse, Nicky got her handbag stolen. I was quite glad to get back to work after that. Anyway, hope you can make it on Saturday—let me know if you want to stay over. Bye!' (pp 105–106)

During this task they either doodled or did not doodle (the control group). This was the independent variable. They were told beforehand they would be tested on the names of people who were attending the party (and not the ones who were not going to be there). This was the 'monitoring' task. They also had an unexpected test, on the names of places mentioned. This was the 'recall' task. The order of these tests was **counterbalanced**, i.e. half the participants were asked to recall the names of party-goers then the places mentioned. The other half recalled the places first, then the names. These two tasks were the measures of the dependent variable (DV) of recall. To operationalise the DVs, plausible mishearings, such as 'Greg' for 'Craig', were counted as correct. Other names that were on the tape but were not party-goers (e.g. John) were scored as false alarms. Other words relating to people, such as 'sister', were ignored. The final score for monitoring was the number of correct names minus false alarms.

Reflections: Think carefully about the difference between 'mishearings' and 'false alarms'. Why was it important that they were treated differently in the collection and analysis of results?

RESEARCH METHODS

Counterbalancing is often used as a control procedure against order effects in a repeated measures design when participants encounter both levels of the IV. Here it is used to control for potential order effects caused by the two different measures of recall (the DV).

The mock telephone message lasted 2.5 minutes and was recorded in a monotonous voice at an average speed of 227 words per minute. It had eight names of people attending a party, and the names of three people and one cat who could not attend. Eight place names were mentioned, as well as irrelevant details.

The participants were given the standardised instructions: 'I am going to play you a tape. I want you to pretend that the speaker is a friend who has telephoned you to invite you to a party. The tape is rather dull but that's okay because I don't want you to remember any of it. Just write down the names of people who will definitely or probably be coming to the party (excluding yourself). Ignore the names of those who can't come. Do not write anything else.' (pages 101–102).

An A4 sheet was given to the participants in the doodling condition, with alternating rows of squares and circles, ten per row. There was also a wide margin on the left for recording the target information. These participants were also given a pencil and asked to shade in the squares and circles while listening to the tape. They were told 'It doesn't matter how neatly or how quickly you do this – it is just something to help relieve the boredom' (page 102). The control participants were given a sheet of lined paper to write their answers on (which they could also have used for doodling).

Each participant listened to the tape at a comfortable volume and wrote down the names as instructed. The experimenter collected the response sheets, then talked to the participants for one minute, including an apology for misleading them about the memory test. They then completed the surprise test of recalling names of places then people or vice versa.

Results

In the doodling condition, the mean number of shaded shapes on the printed sheet was 36.3, with a range of 3–110 and no participants in the control condition doodled spontaneously.

Task type	Measure	Group	
		Control	Doodling
Names (monitored information)	correct	4.3	5.3
	false alarms	0.4	0.3
	memory score	4.0	5.1
Places (incidental information)	correct	2.1	2.6
	false alarms	0.3	0.3
	memory score	1.8	2.4

Table 3.1 Mean recall for doodling and non-doodling groups

RESEARCH METHODS

Three doodlers and four controls suspected a memory test. This suggests that there were **demand characteristics** that made the aim apparent to the participants. However, none said they actively tried to remember information.

Participants in the control group correctly recalled a mean of 7.1 (SD 1.1) of the eight party-goers' names and five people made a false alarm. Participants in the doodling group correctly recalled a mean of 7.8 (SD 0.4) party-goers' names and one person made one false alarm, see Table 3.1. Overall, the doodling participants recalled a mean of 7.5 names and places, 29% more than the mean of 5.8 for the control group. Recall for both monitored and incidental information was better for doodlers than controls, even when the participants who suspected a test were excluded (to eliminate effects due to **demand characteristics**).

Conclusions

Doodling helps concentration on a primary task as the doodling participants performed better than participants just listening to the primary task with no concurrent task. However, because the doodling group were better on both the monitored and incidental information there are two possible explanations. Either the doodlers noticed more of the target words, an effect on attention, or doodling improved memory directly, for example by encouraging deeper information processing. However, without any measure of daydreaming (which could have blocked attention) it is difficult to distinguish between these two explanations. This could have been done by asking participants about daydreaming retrospectively by self-report. Alternatively a simultaneous brain scan could have indicated whether doodling reduced activation of the cortex, which is associated with daydreaming.

ISSUES AND DEBATES

Doodling could be a useful strategy when we have to concentrate and don't want to, for example in an important but boring lecture or when you are waiting to hear the faint sound of a friend's car arriving. By stopping our minds from straying we should be better able to focus on the primary task. This would be an **application to everyday life**.

Reflections: Although doodles have some shape and form they are relatively unplanned, so require little processing themselves but may be sufficient to prevent us from daydreaming.

Conduct an interview or a questionnaire to collect self-report data about the types of doodles that people do and the situations in which they doodle.

Strengths and weaknesses

The main method was a laboratory experiment using an **independent measures** design. This means that it was possible to control **extraneous variables**, for example ensuring the participants were listening at a volume comfortable for them and using a recorded telephone message so that there were no differences in stress on the important words between conditions. It was also **standardised** so that the participants were all equally likely to be bored and therefore to daydream. This was achieved by the monotony of the recording, using a dull, quiet room and asking them to do the experiment when they were expecting to go home. This means the research was more **valid** – they could be sure that the differences in results between conditions were due to doodling or not – and more **reliable**, because all participants were similarly bored. The **operationalisation** of doodling was also standardised, using the doodling sheets, otherwise there may have been individual differences in doodling between participants and some more may not have doodled at all. This also increased **validity**. Nevertheless, there is a risk of **participant variables** confounding the results, as the amount of shapes the individuals shaded differed. However, it was an effective strategy as no participants in the non-doodling condition did doodle (this was discouraged by giving them lined paper) and only one in the doodling condition did not (and they were replaced).

Although the participants were varied in age (18–55 years), so were representative in this respect, they were all members of a recruitment panel and the kinds of people who volunteer for such panels may all be very similar, for example having time to spare or an interest in psychology. This could bias the sample, lowering validity. There was a risk of **demand characteristics** because some of the participants suspected a memory test, but they were roughly equal in each condition and did not actively try to remember, so this is unlikely to have reduced validity. The study collected quantitative data, the number of names and places recalled, which is an objective record of memory. However, it would also have been useful to have asked the participants for self-reports of any daydreaming as this would have helped to explore whether the cause of the difference was attention or memory. It would be important only to ask participants whether they daydreamed, not what they were daydreaming about, as this would be an invasion of their privacy.

This study raised a few ethical issues. The participants were unable to give fully informed consent as they were given an unexpected test on place names. This had the potential to make them distressed if they were unable to remember the names, so could expose them to risk of psychological harm. Although a debrief is not a substitute for good ethical procedures, the experimenters debriefed the participants and apologised for misleading them about the unexpected recall test.

RESEARCH METHODS IN PRACTICE

Individual differences between participants arose. In an **independent measures** design such as this, such differences are a greater problem than they would be in a repeated measures design. This is because they may cause differences in the DV, of memory, that are not caused by the IV but may appear to be so.

3.3 Core study 2: Baron-Cohen et al. (Eyes test)

Baron-Cohen, S., Wheelwright, J., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' Test revised version: A study with normal adults, and adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241–251.

Aim

The main aim of this research was to test whether a group of adults with Asperger Syndrome (AS) or High-functioning Autism (HFA) would be impaired on the revised version of the 'Reading the Mind in the Eyes' task.

The researchers also wanted to test if there was an association between performance on the revised 'Reading the Mind in the Eyes' task and measures of autistic traits, and to investigate if there were sex differences in those without autism on this task. There were five hypotheses:

- Participants with autism will score significantly lower scores on the revised 'Reading the Mind in the Eyes' task than the control group.
- Participants with autism will score significantly higher on the **Autism Spectrum Quotient Test** (AQ) measure.
- Females in the 'normal' groups (Groups 2 and 3) will score higher on the 'Reading the Mind in the Eyes' task than males in those groups.
- Males in the 'normal' group (Group 3) would score higher on the AQ measure than females.
- Scores on the AQ and the 'Reading the Mind in the Eyes' task would be negatively correlated.

KEY TERM

Autism Spectrum Quotient Test (AQ): a self-report questionnaire with scores ranging from 0 to 50. A higher score suggests that the person completing it has more autistic traits.

Background

Autism is a failure to develop particular cognitive processes linked to social interaction that occurs in approximately 1% of the population. Individuals with autism share difficulties in social functioning, communication and coping with change, often alongside unusually narrow interests. The word autism is used as a general term to refer to a range of disorders from HFA and AS to individuals showing severe forms of autism.

In children, autism is characterised by a triad of impairments:

- difficulties with social interaction
- difficulties with verbal and non-verbal communication
- a lack of imaginative play.

Baron-Cohen suggests that people with autism lack or have an underdeveloped cognitive process called a 'theory of mind'. A theory of mind is a cognitive ability enabling us to realise that others have different feelings, beliefs, knowledge and desires from our own. Individuals with autism find it difficult to understand that other people have their own plans, thoughts or points of view. A theory of mind is often linked to empathy. Empathy is the ability to understand the world as another person does, to appreciate their feelings or emotional state separate from our own.

There are several different tests available to measure a person's theory of mind but most of these are designed for children. Baron-Cohen et al. developed a test called the 'Reading the Mind in the Eyes' task to use with adults to test their ability to attribute emotional labels to others. He suggests that this task tests the first stage of theory of mind: assigning an appropriate mental state to another and that the task of 'Reading the Mind in the Eyes' is a good measure of social cognition.

Two key publications in the field of social cognition are the **Diagnostic and Statistical Manual (DSM)** and the **International Classification of Disorders (ICD)**. Autism was first included as a named category in the DSM 3rd edition (1980) when it was called infantile

KEY TERMS

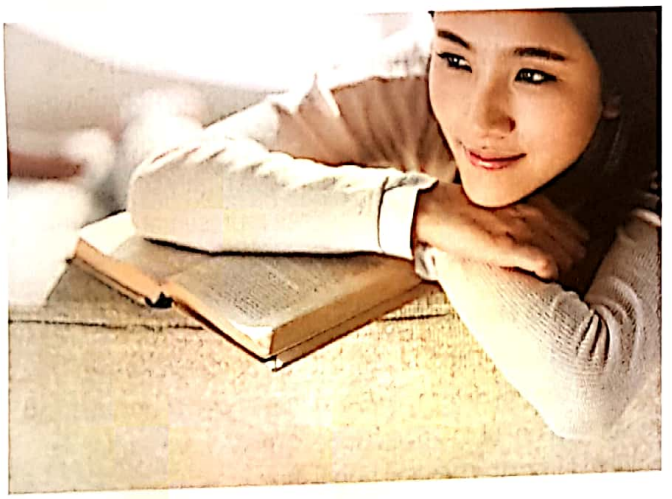
Diagnostic and Statistical Manual (DSM): published by the American Psychiatric Association, it is used as a classification and diagnostic tool by doctors, psychiatrists and psychologists across the globe.

International Classification of Disorders (ICD): published by the World Health Organization (WHO) and although similar to the DSM, it has a wider scope and covers all health-related conditions, not only mental health and psychological conditions.

autism. In the DSM 5th edition (2014), individuals who are diagnosed with autism need to meet two criteria. These are an impairment of social communication and social interaction skills and evidence of restricted, repetitive patterns of behaviour, interests or activities.

When considering the revisions, it was suggested that there was not enough evidence to show a definite distinction between AS and HFA spectrum disorder. Therefore, it was decided to combine both of these terms into a category called 'autism spectrum disorder'.

Using this 'Reading the Mind in the Eyes' task, in 1997, Baron-Cohen et al. conducted an experiment to investigate whether adults with HFA or AS had problems employing a theory of mind. To test this, Baron-Cohen et al. compared a group of individuals who had HFA or AS with some who did not. The participants were shown photographs of eyes and asked to identify the emotion that was being shown from two options. This was the Reading the Mind in the Eyes Task (see Figure 3.5).



3.5 Fantasising or noticing

Baron-Cohen et al. suggested that this test demands similar cognitive processes as having a theory of mind as it requires the participant to empathise with the person in the photograph to infer their emotional state.

From the results in 1997, Baron-Cohen et al. concluded that those high-functioning adults with autism or AS could identify significantly fewer emotions correctly in the 'Reading the Mind in the Eyes' task than 'normal' participants in the **control group**. However, he suggested that there were several practical issues with the 'Reading the Mind in the Eyes' task:

- It was a forced choice question with only two responses and these were always opposites (e.g. Sympathetic or Unsympathetic).
- The small number of examples in the test (25) led to many in the 'normal' group scoring 24 or 25, causing a **ceiling effect**.
- The 25 sets of eyes illustrated both **basic emotions** and **complex emotions** and the former were too easy.
- The emotion in some of the photos used in the original task could be solved by checking the direction to which the person was gazing (e.g. ignoring).
- There was an imbalance of male and female faces.
- Participants might not have understood the words in the 'Reading the Mind in the Eyes' task.

The following 2001 study attempted to revise the 'Reading the Mind in the Eyes' task as a measure of the theory of mind and aims to check the reliability of the results from the previous 1997 study.

KEY TERMS

control group: often used in experiments, this group does not receive the manipulation of the independent variable and can be used for comparison with the experimental group or groups.

ceiling effect: this occurs when a test is too easy and all participants in a condition score the top score. This is problematic as it does not allow the research to differentiate between participants.

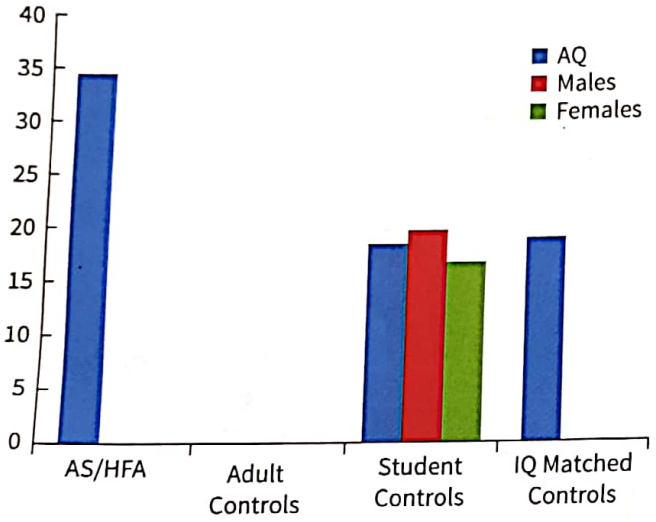
basic emotions: as argued by Ekman (1992), there are six basic emotions that are recognised universally by adults and even very young children developing normally. These emotions, illustrated in Figure 3.6, are happiness, sadness, anger, surprise, fear and disgust.

3.6 Universally recognised emotions

On the AQ task, participants with AS/HFA scored significantly higher than the student comparison and IQ matched comparison groups (see Figure 3.9). There was a significant difference between male and female AQ scores in the student comparison group.

RESEARCH METHODS

The **measure of central tendency** used to calculate the averages used for Figures 3.8 and 3.9 were both **means**. What other measures of central tendency could have been used?



3.9 Average scores on the AQ Test

There was a significant negative correlation between the AQ and Eyes Task scores (–0.53) but no ***correlation*** *between the IQ and Revised Eyes Test scores. This suggests that as a participant's AQ score increases (illustrating higher autistic traits) their ability to correctly identify the correct target word on the Eyes Test decreases.*

Conclusions

The results suggest that the participants with AS or HFA have a deficit in a cognitive process that allows a person to identify emotions in other individuals. This lack of a theory of mind, or ability to attribute emotions to another person, is strongly linked to autism spectrum disorders.

There was evidence of a sex difference between males and females in the comparison groups; with males showing more autistic traits and performing worse on the Eyes Test than females. However, some of these differences were not significant so further research would need to be conducted to clarify if there are differences between sexes and their AQ ability to attribute emotions to others.

It is apparent that the revised Eyes Test used in this study was a more sensitive measure of adult social intelligence than that used in previous studies and will allow future research to discriminate individual differences in a more meaningful way.

Strengths and weaknesses

The main method was a laboratory experiment as all participants completed the task in a standardised way in an artificial setting. This allowed many confounding variables to be controlled for as all participants saw the same set of eyes for exactly the same amount of time. This improves the internal validity of the experiment and allows the research to be repeated to check the reliability of the results. The validity of the Eyes Test has improved on previous versions due to the changes made which have led to normal performance being significantly below the ceiling of the test. As this is a natural/quasi-experiment, it was not possible for Baron-Cohen et al. to randomly allocate participants to the conditions. This introduces a confounding variable, as it could be another factor that is causing the difference between the scores in the different conditions. This research goes some way to attend to this issue by having two different control groups where some were matched on IQ to make the groups as similar as possible.

Although the Eyes Test was vastly improved for this research, there are still several issues that affect the ecological validity of the task. In an everyday situation a person's eyes would not be static and shown for a limited amount of time. Consequently, any attempt to apply the results from this research to an everyday situation will be flawed. Future research might choose to use videos of eyes rather than images to improve the validity of any conclusions.

The experimental sample in this research (AS/HFA participants) is small; therefore when generalising the results from the research we must be aware that the group may not be representative of all individuals who have been diagnosed with AS/HFA.

Reflections: Can we be sure it is really all in the eyes? What else could be important when we are interpreting emotions? To answer this question, think about whether there are any issues with using photographs of eyes rather than a real person in this study. How could this experiment be conducted in a more ecologically valid way? What impact would it have?

SELF-ASSESSMENT QUESTIONS

1 Give **two** examples of how Baron-Cohen et al.'s (2001) study on autism could lack ecological validity.
2 Give **two** problems with the original Eyes Test and for each describe how Baron-Cohen et al. addressed the problems in this study.
3 How could the Eyes Test be further improved?

3.4 Core study 3: Laney et al. (false memory)

Laney, C., Morris, E., Bernstein, D., Wakefield, B., & Loftus, E. (2008). Asparagus, a love story. Healthier eating could be just a false memory away. *Experimental Psychology, 55*(5), 291–300.

Background

Memory is not always a factual recording of an event and can become distorted by other information both during encoding and after the event. There have been many experiments within memory research that have demonstrated that memories can be distorted by information provided following an event. This has even resulted in people believing that an impossible event has taken place. Braun et al. (2002) found that it is possible to implant these rich false memories by convincing participants that they had met Bugs Bunny (Figure 3.10) at Disneyland (impossible as Bugs Bunny is a Warner Bros creation).

3.10 Could you really meet Bugs Bunny at Disneyland?

Could there be behavioural consequences of false memories? Other studies by Laney et al. (Bernstein et al., 2005) have shown that false memories of sickness following the eating of pickles or eggs resulted in a number of participants changing their willingness to eat the foods when asked about their food preferences. If false beliefs about illness could result in people avoiding foods, could a positive false memory encourage people to eat a food?

3.11 Could a false memory actually change your preference for a food?

This research investigated the impact that implanting positive false beliefs and memories in people

RESEARCH METHODS

Questionnaires can contain different types of questions: **open questions** which allow participants to express themselves in their own words and **closed questions** which differ a limited number of fixed choices. Look at the examples of questions below. Are these open questions, closed questions or a mixture?

Food History Inventory (FHI) where respondents rated 24 items on a scale of 1 (definitely did not happen) to 8 (definitely did happen) as to their food experiences before the age of 10 (e.g. 'Loved asparagus the first time you tried it').

Restaurant Questionnaire (RQ) which assessed the respondents' desire to eat each of 32 separate dishes (e.g. sautéed asparagus spears). This questionnaire was designed to look like a menu with five 'courses'. Respondents were asked to imagine they were out for a special dinner and to rate, regardless of price, how likely they would be to order each food on a scale of 1 (definitely no) to 8 (definitely yes).

Food Preferences Questionnaire (FPQ) was a 62 item inventory of items of food (e.g. asparagus) that respondents had to rate on a 1 (definitely don't like to eat, for whatever reason) to 8 (definitely like to eat).

Food Costs Questionnaire (FCQ) was a list of 21 different food items (e.g. a pound of asparagus) with multiple choice answers where respondents had to circle the price they would be willing to pay for each, including a 'would never buy' option. For asparagus, the price options were $1.90, $2.50, $3.20, $3.80, $4.40, $5.00 and $5.70.

Memory or Belief? Questionnaire (MBQ) – respondents were asked to indicate whether they had a memory of an experience with three items from the FHI, including, for the 'love' condition, the critical item of asparagus. The choices were that they had specific memory of the event occurring, a belief that the event had occurred (but lacked specific memory), or were positive that the event had not occurred.

about their liking of eating asparagus as a child would have on their food preferences later. Laney et al. expected to be able to implant memories of loving asparagus the first time it was tried from previous research, but wanted to investigate if these false memories would lead to positive consequences for their participants.

The research consisted of two experiments. The first aimed to see if false feedback about a liking of a food could cause a false memory and change a participant's eating behaviours. As well as checking the reliability of the findings in the first experiment, the second experiment wanted to examine the underlying cognitive mechanisms of the false memory consequence effect.

Experiment 1: Aim

The aim was to investigate whether giving false feedback suggesting that a participant had loved to eat asparagus as a child, would generate a false belief or memory of experiences linked to eating and enjoying asparagus.

Experiment 1: Method

Research method and design

The experiment was a laboratory experiment as the environment the participants were tested in was very artificial and unlike an everyday event, although some small attempts to make one questionnaire look like a menu were taken. Participants were tested in laboratory settings in groups of up to eight people.

The independent variable was whether a participant had the false belief that they had enjoyed asparagus as a child embedded during the second part of the experiment. These participants were compared with a (control group) of participants who received no false belief.

The dependent variable was measured through the use of five questionnaires (see Research Methods box). A questionnaire is a self-report measure where participants respond through writing their answers down.

Sample

This experiment is an example of an independent groups design as participants only took part in one of the two conditions: the 'love' condition or the control group.

All 128 of the participants were undergraduate students at the University of California who received course credit for their time. The sample consisted of 99 females and 29 males who had a mean age of 20.8 years old. The participants were randomly assigned to either the 'love' condition (63) or the control group (65).

Procedure

Participants arrived at the laboratory in groups of up to eight and were told that they were going to take part in a study of 'food preferences and personality'. This deception was necessary to limit **demand characteristics** through awareness of the true aim.

During the first session (week one), all participants were treated identically. Participants first completed the FHI and RQ. The researchers also asked participants to complete three other questionnaires. These were to distract them from the true aim of the study. These other questionnaires included a personality measure, a social desirability scale and an eating habits questionnaire.

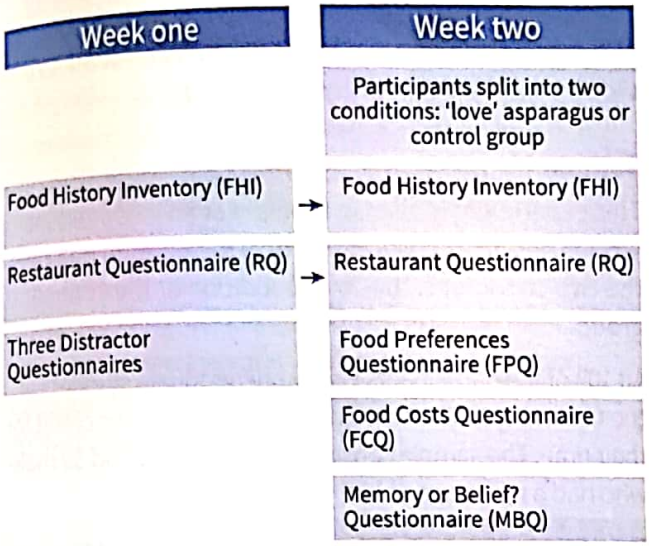
Approximately one week later participants were invited back to the laboratory. At this point participants were randomly allocated to the 'love asparagus' condition or the control group. All participants were told that their responses from the first week had been processed by a computer which had generated a profile of their early childhood experiences with food and were given a report that included:

> *'As a young child, you disliked spinach, you enjoyed fried foods, and you felt happy when a classmate brought sweets to school.'*

The critical item, 'you loved to eat cooked asparagus', was embedded in the third position of the profile for participants in the 'love' condition. Those in the control group only received the three filler items. Participants were then asked questions about this fake profile to ensure that they had processed the feedback. These questions included:

- Imagine the setting in which this experience might have happened. Where were you? Who was with you?
- On a scale of 1 (not at all) to 9 (very much), to what extent did this experience affect your adult personality?

Following this, participants completed the FHI and the RQ a second time to measure any changes in responses from before the implanting of the false belief. Further to these, participants also completed three further questionnaires: FPQ, FCQ and MBQ.



3.12 Overview of the questionnaires completed by participants at week one and week two

When participants had completed these questionnaires (Figure 3.12) they were fully debriefed by the researchers and told the true nature of the experiment.

Experiment 1: Results

The two key issues the researchers wanted to investigate were whether subjects formed false asparagus-related beliefs and whether these beliefs have consequences.

When asked a second time as part of the FHI if a participant loved asparagus the first time they tried it, participants in the 'love' condition's average (mean) response rose by 2.6 points following the false feedback from the researchers. The responses from those in the control condition only increased by 0.2 points in comparison. This was a statistically significant difference between the conditions.

	Week one	Week two
Love ($n = 46$)	1.5	4.1
Control ($n = 51$)	1.5	1.7

Table 3.4 Mean ratings of the critical item ('Loved asparagus the first time you tried it') on Food History Inventory

Thirty-one participants were excluded from this analysis as they initially believed they loved asparagus the first time, or scored greater than five on the FHI on the first occasion. This left 97 participants to be included in the analysis.

Memories or beliefs?

For the purpose of this study, memories are the ability to recall specific structured events with some details; the participant 'remembers' an experience. Beliefs are less detailed and not tied to a specific time or place; the participant 'knows' it happened but cannot go into specific detail.

When asked if they had specific memories or a belief that they loved asparagus the first time they tried it, there was a difference between the 'love' condition and the control group. However this was not a statistically significant difference.

	Memory or belief			Not the case
	Memory (M)	Belief (B)	M or B	
Love (46)	22% (10)	35% (16)	57% (26)	43% (20)
Control (51)	12% (6)	27% (14)	39% (20)	61% (31)

Table 3.5 Responses to the question asking if a participant had a specific memory, belief, or it was not the case that they loved asparagus the first time they tried it (MBQ)

This suggests that those participants who were told that they loved asparagus when they first tried it had a greater chance of generating a false memory or belief to substantiate this false memory.

Believers vs nonbelievers

To be classified as a believer, participants had to meet the following three criteria:

- given a low rating on the FHI when initially asked if they loved asparagus (week one)
- increased their rating on the FHI when asked if they loved asparagus on week two
- given positive 'memory' or 'belief' response on the MBQ.

Forty-eight per cent (22) of participants in the 'love' condition met the criteria to be labelled believers and further analysis compared their scores with the other 'non-believers' where the impact of the implanted memory had a lesser effect. The ratings of these 22 believers increased an average of

4.5 points from week one to week two on their FHI item. Nonbelievers increased an average of just 0.9 points. Of the 22 participants classified as believers, 10 claimed to have an asparagus 'memory' and 12 claimed a 'belief' on the MBQ.

To assess the consequence of false beliefs, the believers were compared with those in the control group.

On the RQ, believers reported more desire to eat the critical asparagus item than those in the control group. On the FPQ, in comparison to the control group (mean 3.84), believers (mean 6.14) reported liking asparagus significantly more. Finally, on the FCQ, believers were willing to pay significantly more for asparagus than those in the control group with over a quarter (14) of those in the control group stating that they would never buy asparagus. None of the believers selected the never buy response.

Experiment 1: Conclusions

Participants can be led to develop positively framed false beliefs and these false beliefs can have a consequence on behaviour and food preferences. Participants who had the false belief implanted increased their rating of their love of asparagus and these beliefs had further impacts on how much they would be willing to spend on asparagus, greater intention to eat asparagus in the future, and a greater preference for it.

In the second experiment the researchers wanted to explore why these false beliefs about having loved a healthy food the first time one tried it led to increased liking of that food.

Experiment 2: Aim

The aim of the second experiment was to examine the possible underlying mechanisms of the false memory consequence effect by exploring if, after the false love of asparagus manipulation, the very sight of asparagus was more appealing to participants. A secondary aim was to replicate and extend the results of the first experiment to check the reliability of the findings.

Experiment 2: Method

Research method and design

The independent variable was whether the participant had the false belief, 'you loved asparagus', embedded. These participants were compared with a control group of participants who received no false belief.

The dependent variable was measured through the use of four questionnaires and the participant's feedback to a slideshow of 20 pictures of common foods.

Sample

This experiment is also an example of an independent groups design as participants only took part in one of the two conditions: the 'love' condition or the control group.

All 103 of the participants were undergraduate students at the University of Washington who received course credit for their time. The sample consisted of 64 females and 39 males who had a mean age of 19.9 years.

The participants were randomly assigned to either the 'love' condition (58) or the control group (45).

RESEARCH METHODS

A slideshow was used to present the photographs, which helped to standardise the procedure. This means that differences between conditions were more likely to be due to the IV than to differences in the way the participants were treated. Imagine doing the same test in a classroom using actual photographs that you flip over in a stack in front of the participant. Even if you use a stopwatch to time yourself, you won't be as accurate and the movement of the photographs could distract the participant.

Procedure

Participants arrived at the laboratory and were told that they were going to complete several questionnaires to generate a personal profile of eating experiences based on their responses. There was no deception or cover story for this experiment. All participants first completed the FHI, the RQ and the FPQ. Participants also completed two distractor questionnaires: a personality measure and a social desirability scale.

Similar to the first experiment, approximately one week later participants were invited back to the laboratory. At this point participants were randomly allocated to the 'love' asparagus condition or the control group. All participants were told that their responses from the first week had been processed by a computer which had generated a profile of their expected early childhood experiences with food.

Participants in the 'love' condition were given a profile that contained the critical phrase 'you loved asparagus the first time you ate it' in the third position. After reading the profile, participants then completed an elaboration exercise in which they were required to give details about their memory of eating asparagus, or if they had no memory of it, what might have happened. Those allocated to the control group were told nothing about asparagus and did not complete the elaboration exercise. All participants were then asked: what is the most important childhood, food-related event in your life that your food profile did not report?

3.13 Pictures were rated by participants

A slideshow of 20 colour photographs of common foods were displayed for 30 seconds each to all participants (Figure 3.13). Participants were asked four questions about each slide. On a scale of 1 (not at all) to 8 (very much):

- how appetising they found the food depicted in the photo
- how disgusting they found the food depicted in the photo
- the artistic quality of the photo
- whether the photo was taken by a novice, amateur or expert photographer.

Participants finally completed the RQ, FPQ, and the FHI for a second time, and the same MBQ as in Experiment One (Figure 3.14). When all questionnaires were completed, the participants were fully debriefed.

Week one	Week two
	Participants split into two conditions: 'love' asparagus or control group.
Food History Inventory	→ Food History Inventory
Restaurant Questionnaire	→ Restaurant Questionnaire
Food Preferences Questionnaire	→ Food Preferences Questionnaire
Two Distractor Questionnaires	Common Foods Slideshow
	Memory or Belief Questionnaire

3.14 Overview of the questionnaires completed by participants at week one and week two

Experiment 2: Results

Food History Inventory

As in Experiment 1, the 'love' and control groups rated their liking of asparagus similarly before the manipulation, but differently after they had received their profile distorting their memory. This was a statistically significant difference between the conditions.

	Week one	Week two
Love (n = 40)	1.7	4.2
Control (n = 33)	1.5	2.5

Table 3.6 Mean ratings of the critical item ('Loved asparagus the first time you tried it') on Food History Inventory

Excluded from this analysis were 30 participants (18 'love' group subjects and 12 controls) who were reasonably sure that they had loved asparagus the first time they tried it before the manipulation (with a rating of five or higher on the FHI).

Memories or beliefs?

Similar to the first experiment, the results from the MBQ suggested that those participants who were told that they loved asparagus had a greater chance of generating a false memory or belief to substantiate this false memory. However this was not a statistically significant difference.

	Memory or belief			Not the case
	Memory (M)	Belief (B)	M or B	
Love (40)	28% (11)	28% (11)	57% (22)	45% (18)
Control (32)	6% (2)	38% (12)	39% (14)	56% (18)

Table 3.7 Responses to the question asking if a participant had a specific memory, belief, or it was not the case that they loved asparagus the first time they tried it (MBQ)

Believers vs nonbelievers

Participants were separated into believers or nonbelievers based on the same criteria as in the first experiment. Forty participants in the 'love' condition met the criteria to be labelled believers. The believers were then compared with those participants in the control group.

On the RQ, neither the believers nor the control group reported an increased desire to eat the critical asparagus item when comparing the two weeks. On the FPQ, in comparison to the control group, believers reported a significantly greater desire to eat asparagus. Finally, on the photograph ratings, believers rated the asparagus photo as more appetising than those in the control group (5.10 versus 4.00), and as less disgusting (1.81 versus 3.24).

Experiment 2: Conclusions

Participants can be given positive false food beliefs and these beliefs have consequences on behaviours and attitudes towards foods. Those participants who believed the false feedback were more likely than those in the control group to rate a photograph of asparagus as more appetising and less disgusting.

The photograph measure provides a step towards understanding the cognitive mechanisms associated with false memories as the false memory primed the participant to process the images of asparagus more positively. This positive response is interpreted as familiarity and the participants misattribute it to a childhood experiences (I did love asparagus the first time I tried it) and consequently, an adult preference (I love asparagus).

Across the two experiments it was shown that participants could have false beliefs implanted about whether they had previously had a specific positive experience with asparagus and that this belief had consequences on their attitudes and even memories about that food.

ISSUES AND DEBATES

This study has a clear application to everyday life. If implanting false memories could encourage people to eat more healthily, for example because they began to like fruit and vegetables more, it could contribute to solving problems like obesity. This would be a useful strategy to help people who are already trying to diet but have developed bad eating habits like snacking on crisps and sweets. However, being aware that they were having false memories implanted may prevent the effect from working outside the laboratory.

Strengths and weaknesses

The main method was a laboratory experiment as all of the participants were tested in a **standardised** environment and given the same questionnaires in each condition. The only difference between the experiences of the participants in this experiment was the introduction of the critical 'you loved asparagus' comment on the generated profile for each participant. Each questionnaire was standardised, and further questionnaires were used as distractors to prevent the participants from working out the true aim of the experiment and controlling demand characteristics. This means that the research has more internal validity as we can be more confident that the independent variable is the only variable having an impact on the dependent variable.

The use of questionnaires as a method to operationalise the dependent variable allowed Laney et al. to further standardise the collection of the data within the experiment. Questionnaires allow researchers to collect quantitative data quickly from participants that can be easily analysed and used to compare results across two conditions. However, it is worth noting that we do not know for certain whether these effects will translate to actual eating behaviours. Completing a questionnaire may not involve the same processes as choosing to eat (or not eat) a specific food in a restaurant setting, therefore we could argue that there is little ecological validity in this research.

As this study was conducted in a short space of time and the participants were not followed up, we are not able to ascertain how long the effects of the false memories last. The participants completed the questionnaires within a few minutes of receiving false feedback so there is no way of knowing how long lasting the impact will be. More research of a longitudinal design will be needed to research the durability of any attitude and behavioural consequences of the false memories.

The sample in this research consists of university students and introduces participant variables that could distort the outcomes of the research reducing the validity of the research. Students might react differently to the false information and be more impressionable than an adult or child.

Reflections: Think about a food you have a strong feeling about (either like or dislike) – can you remember why you dislike it? Could this be a false memory? Talk to your parents about it. Could our memories of past events really change your liking or disliking of a food? How could we use the findings from this research to change a person's eating habits?

SELF-ASSESSMENT QUESTIONS

4 Jaina is planning a study about false memories. She wants to compare false memories about emotional events and false memories about non-emotional events.
   a Suggest how Jaina could operationally define the IV in her study.
   b Suggest one possible ethical problem with Jaina's study.

5 Patients who receive chemotherapy in hospital are often very sick and this makes them lose their appetite. Using the study by Laney et al., suggest how a procedure to help patients to believe foods taste nice could be developed.

3.5 Issues, debates and approaches

The application of psychology to everyday life

When we are trying to focus, such as when listening to a lecture, allowing ourselves to doodle might be advantageous. However, any attempt to deliberately 'draw' something specific is likely to be counterproductive as the drawing itself might become the primary task, distracting us from the lecture rather than allowing us to concentrate better.

The eyes provide a lot of information when we are attributing an emotion to an individual. It might be possible to develop a programme to help teach individuals diagnosed with AS/HFA to help them develop skills of interpreting emotions. The Eyes test could also be developed further to help aid initial diagnosis of individuals who could be signposted to appropriate clinical staff to investigate if there is an underlying autistic disorder.

Laney et al. demonstrate that it is possible to impact some people's attitudes towards asparagus by simply giving a small amount of false information. This could be used to help people change their diets and become healthier or help people change their attitudes towards healthier foods such as salad and vegetables.

Individual and situational explanations

As doodling affected recall, this means it has a situational effect information processing. Furthermore, as Andrade deliberately ensured that her participants would be bored so that they were more likely to doodle, this implies that there are situational causes for doodling itself. However, there were very large differences between the doodles of participants initially allocated to the doodling group (from 0 to 110 shaded shapes) and we know that people who doodle do not all doodle in the same way. This means that there are also individual causes behind doodling behaviour.

Those individuals who had a diagnosis of AS/HFA performed significantly worse on the Eyes test than 'normal' individuals, suggesting that the ability to read emotions in the eyes is an individual skill that is developed, rather than being the result of the external environment. The environment was standardised across those participants in both the AS/HFA and 'normal' conditions, providing further support for the individual explanation.

The effect of the false memory embedded by telling one group of participants that they loved asparagus as a child means that the situation could affect an individual's attitude towards the food. As Laney et al. were careful to control that the information presented to the participants was the same, other than the statement about a love of asparagus, this suggests the information from the situation had a consequence on later behaviour. There were individual differences among participants however, with some participants in the control group liking asparagus and others in the 'love' asparagus condition failing to believe that they had experienced this love as a child.

Nature vs nurture

This is a long-running debate, which considers whether we are the way we are because of nature (inherited and genetic) or nurture (experiences and influences after conception). There is no known single cause for autism and there are both generic and environmental arguments. Several different genes appear to be involved in autistic spectrum disorder, with some of these being inherited and others happening spontaneously. Researchers are currently investigating if environmental factors such as viral infections, pollutants or issues during pregnancy could have a role in triggering the disorder.

Children as participants

The research in this area does not focus on children as participants, so it might be useful to consider how similar research could be conducted on younger individuals. Baron-Cohen et al.'s research was conducted on adult male participants using pictures of adult eyes and relatively sophisticated words to describe the emotion displayed. This would not be appropriate for children and alternative methods should be considered to investigate similar aims in children. Baron-Cohen has devised a 'Theory of Mind' test for children called the 'Sally and Anne' test that involves asking children questions following a short scene that is acted out with dolls.

Summary

Andrade's study tested whether doodling could improve concentration on and memory of a conversation. Participants in the doodling condition *remembered* more of the people's names they had been asked to recall and *attended to* the message better, as they recalled more in a surprise test of the place names mentioned. This laboratory experiment was well controlled with a recorded stimulus message and specific shapes to colour in for the doodlers. The data from the words recalled was quantitative and objective. However, it would also have been useful to have had qualitative data about whether the participants daydreamed as this would have helped to distinguish between two possible reasons for the improved memory: deeper processing or better attention.

Baron-Cohen et al.'s study investigated how a lack of a 'theory of mind' in adults on the autistic spectrum relates to problems recognising emotions. The autistic spectrum group's scores on the Eyes test were worse than the control group on the Eyes test even though their IQ was no different. There was also a negative correlation between the revised Eyes test score and Autistic Spectrum Quotient. The test itself was valid because the eyes were shown for a fixed amount of time, although this could also be a weakness as emotions are usually detected on live faces, which move. The findings suggest that the revised Eyes test is better at detecting individual differences in social sensitivity than the previous version, i.e. it is more valid.

Laney et al.'s study explored whether 'false memories' could alter memories and beliefs about eating asparagus. After a suggestion that they loved to eat asparagus as children, participants were more confident that they had loved asparagus the first time they had tried it. This false belief also increased their liking of asparagus, desire to eat it and willingness to pay more for it. This may have been because the false memory made the asparagus look more appealing. Standardisation of the profiles and questionnaires meant that the comparison between the false memory and control groups focused on the 'you loved asparagus' comment, making the study valid. However, the effects may only be short-term as there was no follow up. The findings suggest that adults can be led to believe they had positive food-related experiences as children, which can have healthy consequences.

Exam-style questions

1 Danvir is planning a laboratory experiment about doodling but is worried that the findings may not generalise well to the effects of real-world doodling.

a Suggest **one** reason why the results may not generalise to real-world doodling. **[2 marks]**

b Design a natural experiment to test **one** factor that could affect doodling. **[10 marks]**

2 Suggest why it was important to the study by Andrade that the telephone message used was boring. **[2 marks]**

3 In the original Eyes test used by Baron-Cohen et al., there was an imbalance of male and female faces. Explain why this could have been a problem. **[2 marks]**

4 Identify the different types of intelligence measured in the study by Baron-Cohen et al. (Eyes test). **[2 marks]**

5 Explain what the results of the Baron-Cohen et al. study show about different types of intelligence in people on the autistic spectrum. **[4 marks]**

6 Describe **one** aim of the study by Laney et al. (2008). **[2 marks]**

7 Explain why Laney et al. concluded that false beliefs could affect behaviour. **[2 marks]**

Chapter 4
Learning approach

Introduction

The aim of this chapter is to introduce you to the learning approach to psychology and to explore three studies from this approach. They are:

- **Bandura et al. (aggression)** which is based on social learning theory and looks at the effect on children's behaviour of seeing an adult behaving aggressively
- **Saavedra and Silverman (button phobia)** which is a case study of a young boy with a phobia of buttons and the use of classical conditioning to help reduce his fear and disgust
- **Pepperberg (parrot learning)** which explores the comprehension of object categories by a parrot which was trained through social learning and operant conditioning.

These three studies serve to illustrate the main assumptions of the learning approach which are that:

- conditioning helps to explain changes in behaviour
- social learning helps to explain changes in behaviour.

Throughout the chapter you will find out about examples of different types of learning processes and see how these operate with people and with animals. In each core study you will learn about the background to the research, the way it was conducted, the results and the conclusions. From this you will be able to see how the studies, and the learning approach in general, illustrate a range of aspects from the research methods topic. In addition, you will discover how these ideas can be used to illustrate a range of issues and debates.

Live to learn or learn to live?

Think about what you have done today. Apart from basic biological activities, like breathing or blinking, almost everything you have done you will have learned. You are reading this, so you have learned to read, maybe in more than one language. Hopefully you've eaten a meal and maybe travelled to class. How did you learn how to eat or how to get to your classroom?

We can learn through different mechanisms and most of these can also be seen in animals. In some respects, however, our learning is different. Learning means new, permanent changes in behaviour following experience. If you have a pet, or regularly see wild animals, how do you think your learning is similar to and different from theirs? You might have thought of two different kinds of answers. The *way* we learn might differ and *what* we can learn might differ. Both of these are good answers.

When you think about 'learning' the first thing that probably occurs to you is trying to learn your work. In fact this is a memory task. We use 'learning' to refer to the acquisition of new behaviours. Kimble (1961) defined learning as 'a relatively permanent change in behavioural potential which accompanies experience.'

Looking carefully at this statement, we can see that it has three parts:

- learning results in the acquisition of new responses (the relatively permanent change)
- learning may occur without new behaviours necessarily being demonstrated (the behavioural *potential*)
- the environment governs learning (by providing experiences).

In this chapter we will be looking at the changes that occur in learning, how we can measure changes in behavioural potential and how the environment contributes to learning. We will consider repeated exposure to stimuli, rewards and role models, each of which is a feature of the environment.

Looking back over your childhood, think about three different behaviours that you can definitely say you learned, rather than their being purely the product of maturation. Consider how you might have learned them. Did you discover them for yourself or did something or someone in the environment help you? If something or someone was involved, think about what it was that enabled you to learn. Now think about someone quite old. What have they learned recently? Maybe they are developing the ability to use a new piece of technology, or are being told about new games, music or friendships by younger family members. We continue to learn throughout our lifetimes.

Finally, think critically about yourself. Is everything you have learned positive? Have you acquired any bad habits, do you know any words you would not repeat in polite company? The answer to this is likely to be 'yes', even though you probably made no effort to learn these things. This suggests that although some learning involves effort, sometimes we can learn simply by being in a certain environment – and that not all learning is a good thing.

Reflections: Think of something you have learned, such as the meaning of a word. Is it new and permanent? Had you acquired it before actually demonstrating that knowledge? What factor(s) in the environment were responsible for that learning?

4.1 Core study 1: Bandura et al. (aggression)

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63*(3), 575–582.

Aim

The aim was to investigate whether a child would learn aggression by observing a model and would reproduce this behaviour in the absence of the model, and whether the sex of the role model was important. Specifically, there were four hypotheses:

- Observed aggressive behaviour will be imitated, so children seeing aggressive models will be more aggressive than those seeing a non-aggressive model or no model.
- Observed non-aggressive behaviour will be imitated, so children seeing non-aggressive models will be less aggressive than those seeing no model.
- Children are more likely to copy a same-sex model.
- Boys will be more likely to copy aggression than girls.

ISSUES AND DEBATES

nature versus nurture: the importance of the environment, or nurture, can be seen in the role of adults as models for children in the rewards and punishments they give to them.

Background

Children copy adults. This could be because the immediate social setting makes the child imitate what he or she is watching (Figure 4.1). This is just 'facilitation' of behaviour, making it more likely that the child will do what others are doing around them. Alternatively, the observation of a behaviour could lead the child to acquire a new response that he or she could reproduce independently. If this is the case, the new behaviour should generalise to new settings and so would be produced in the absence of an adult model. If this **imitative learning** occurred, it could arise in response to observing either aggressive or non-aggressive behaviour. So whereas watching an aggressive model should lead to more aggressive behaviours being demonstrated, observing a non-aggressive model should lead to more non-aggressive behaviour being produced, i.e. even less aggressive behaviour than normal.

KEY TERM

imitative (social) learning: the learning of a new behaviour which is observed in a role model and imitated later in the absence of that model.

Children are also differentially rewarded for their copying. In general (at least in the mid-twentieth century when this study took place) boys were rewarded for behaviours considered to be sex-appropriate and punished for inappropriate ones, such as cooking or 'playing mother'. Similarly for girls, rewards and punishments would be applied to discourage sex-inappropriate behaviours. This, Bandura et al. suggested, would lead to two kinds of differences. Firstly, boys and girls should be more likely to imitate same-sex models and secondly, they should differ in the readiness with which they imitate aggression, with boys doing so more readily as this is seen as a more masculine-type behaviour.

4.1 A child imitates seen behaviour

Method

Research method and design

This was a laboratory experiment; the environment was not the normal place where the children played and the situation was controlled. The design of the experiment was that of independent measures as different children were used in each of the levels of the independent variables (IVs) (although these children were **matched** for aggression in threes). There were three IVs:

- *model type*: whether the child saw an aggressive model, non-aggressive model or no model
- *model gender*: same gender as child (boys watching a male model and girls watching a female model) or different gender (boys watching a female model and girls watching a male model)
- *learner gender*: whether the child was a boy or a girl.

RESEARCH METHODS

Matched participants design: the participants were divided into threes, all with very similar initial aggression levels. One of each of these individuals was placed into each of the three different conditions of model type (aggressive or non-aggressive model and control).

The dependent variable (DV) was the learning the child displayed. This was measured through a **controlled observation** of the children and measures of aggressive behaviour were recorded.

Sample

Seventy-two children aged three to six years (36 boys and 36 girls) were obtained from Stanford University nursery school.

Procedure

Prior to the experimental part of the study, the children were observed in their nursery school by the experimenter and a teacher who knew them well. They were rated on four five-point scales measuring physical aggression, verbal aggression, aggression to inanimate objects and aggression inhibition (anxiety). They were then assigned to three groups, ensuring that the aggression levels of the children in each group were matched. Of the 51 children rated by both observers (the rest were rated by only one observer), similar ratings were generally produced. Their ratings were compared as a measure of **'inter-rater reliability'**, which showed a high correlation between the observers, of $r = 0.89$.

RESEARCH METHODS

Inter-rater reliability is the extent to which two researchers rate the same activity that they have observed, heard in an interview, etc., in the same way. This is judged using a correlation (an '*r*' value) between the two ratings, which will be high (close to 1) if they are reliable.

Twelve boys and 12 girls were allocated to control groups who saw no model. The remaining children were divided equally by sex between aggressive and non-aggressive model groups and within those, between same and opposite-sex models.

The experimental procedure started with all participants being deliberately mildly annoyed. This was done for two reasons:

- *because watching aggression may reduce the production of aggression by the observer (even if it has been learned) and it was necessary to see evidence of learning (Figure 4.2)*
- to ensure that even the non-aggressive condition and control participants would be likely to express aggression, so that any *reduction* in that tendency could be measured.

4.2 Observing and imitating aggressive behaviours with a Bobo doll

Each individual child was shown to a room with attractive toys such as a fire engine and a baby crib but after about two minutes of play, they were told that these were the best toys and were to be kept for other children. The experimenter and child then moved to the observation room, where the experimenter showed the child to a table and chair in their 'play area', where they were shown how to make potato prints and sticker pictures – activities previously identified as interesting for children. The opposite corner of the room also contained a table and chair, a Tinkertoy set, a mallet and a five foot (152 cm) Bobo doll – an inflatable clown-like doll which bounced back when hit (Figure 4.2). This is where the model sat, in those conditions where there was one. The experimenter remained in the room so that the child would not refuse to be alone or try to leave early but they appeared to be working quietly at their desk.

The three groups were then treated differently. In the non-aggressive condition, the model assembled the Tinkertoys (a wooden building kit) for ten minutes. In the aggressive condition this lasted only one minute after which the model attacked the Bobo doll. The doll was laid on its side, sat on and punched in the nose, picked up and hit on the head with a mallet, tossed up in the air and kicked. This sequence was performed three times over nine minutes accompanied by aggressive comments such as 'Kick him' and two non-aggressive comments such as 'He sure is a tough fella'. Of children in the model groups, half saw a

same-sex model, the others saw a model of the opposite sex. A **control group** did not see any model, and therefore saw no aggression.

RESEARCH METHODS

The non-aggressive model group might appear to be a control group because the key factor of 'aggression' is missing. However, the important aspect is the presence of a model – and there is one here. So the real control group is where the IV is absent, i.e. where there is no model at all.

A test of the child's aggression then followed in which the child was observed for 20 minutes using a one-way mirror. For the aggressive model group, this was a test of delayed imitation. This experimental room contained a three foot (92 cm) Bobo doll, a mallet and peg board, two dart guns and a tether ball with a face painted on it which hung from the ceiling. It also contained some non-aggressive toys, including a tea set, crayons and colouring paper, a ball, two dolls, three bears, cars and trucks, and plastic farm animals. These toys were always presented in the same order.

The children's behaviours were observed in five second intervals (240 response units per child). There were three 'response measures' of the children's *imitation*, with a range of possible activities in each:

- *Imitation of physical aggression*: striking the Bobo doll with the mallet, sitting on the doll and punching it in the nose, kicking the doll, and tossing it in the air.
- *Imitative verbal aggression*: repetition of the phrases, 'Sock him', 'Hit him down', 'Kick him', 'Throw him in the air' or 'Pow'.
- *Imitative non-aggressive verbal responses*: repetition of 'He keeps coming back for more' or 'He sure is a tough fella'.

Reflections: Why might Bandura et al. have chosen to record in five second intervals, rather than one second or one minute intervals?

Partially imitative aggression was scored if the child imitated these behaviours incompletely. The two behaviours here were:

- *mallet aggression*: striking objects other than the Bobo doll aggressively with the mallet
- *sits on Bobo doll*: laying the Bobo doll on its side and sitting on it, without attacking it.

Two further categories were:

- *aggressive gun play*: shooting darts or aiming a gun and firing imaginary shots at objects in the room.
- *non-imitative physical and verbal aggression*: physically aggressive acts directed toward objects other than the Bobo doll and any hostile remarks except for those in the verbal imitation category (e.g. 'Shoot the Bobo', 'Cut him', 'Stupid ball', 'Horses fighting, biting' 'Knock over people'.

Finally, behaviour units were also counted for *non-aggressive play* and sitting quietly *not playing* at all, and records were kept of the children's remarks about the situation.

One male scored all the children's behaviours and, except for those conditions in which he was the model, he was unaware of which condition the child had been in (although this was typically obvious in the case of the aggressive model children as they performed the very specific behaviours exhibited by the model). To test his reliability, a second scorer independently rated the behaviour of half of the children and the reliability was high, around $r = 0.9$ for different categories of behaviour.

Results

Children exposed to aggressive models imitated their exact behaviours and were significantly more aggressive, both physically and verbally, than those children in the non-aggressive model or control groups. These children also imitated the model's non-aggressive verbal responses. This effect was greater for boys than girls although boys were more likely to imitate physical aggression and girls more likely to imitate verbal aggression (though not significantly so). Boys were also more likely to imitate a same-sex model

as, to a lesser extent, were girls. The mean aggression scores can be seen in Table 4.1. Some interesting points are that the mean for imitative physical aggression for male subjects (25.8) is much higher than that for female subjects (7.2). This indicates that the boys imitated the physical aggression of a male model more than the girls. However, with a female model, girls imitated less (5.5) than with the male model.

Children seeing a non-aggressive model were much less likely than either the aggressive model group or controls to exhibit mallet aggression, and this pattern was especially apparent for girls. Although the aggressive model did not appear to affect levels of gun play or punching the Bobo doll, non-imitative physical and verbal aggression other than these activities were higher following exposure to an aggressive model compared to the other two conditions.

Reflections: Consider the data for the mean aggressive behaviours and the non-imitative verbal responses. Which are qualitative and which are quantitative data?

There were also differences in non-aggressive play. Girls played more with dolls, tea sets and colouring and boys engaged in more exploratory play and gun play. There were no gender differences in play with farm animals, cars or the tether ball. Both boys and girls seeing the non-aggressive model engaged in more non-aggressive play with dolls than either of the other groups, and spent more than twice as much time sitting quietly, not playing.

Response category	Experimental groups				Control groups
	Aggressive		Non-aggressive		
	Female model	Male model	Female model	Male model	
Imitative physical aggression					
Female subjects	5.5	7.2	2.5	0.0	1.2
Male subjects	12.4	25.8	0.2	1.5	2.0
Imitative verbal aggression					
Female subjects	13.7	2.0	0.3	0.0	0.7
Male subjects	4.3	12.7	1.1	0.0	1.7
Mallet aggression					
Female subjects	17.2	18.7	0.5	0.5	13.1
Male subjects	15.5	28.8	18.7	6.7	13.5
Punches Bobo doll					
Female subjects	6.3	16.5	5.8	4.3	11.7
Male subjects	18.9	11.9	15.6	14.8	15.7
Non-imitative aggression					
Female subjects	21.3	8.4	7.2	1.4	6.1
Male subjects	16.2	36.7	26.1	22.3	24.6
Aggressive gun play					
Female subjects	1.8	4.5	2.6	2.5	3.7
Male subjects	7.3	15.9	8.9	16.7	14.3

Table 4.1 Mean aggression scores from Bandura et al.'s study

experimenter effect.

ISSUES AND DEBATES

In the nature versus nurture debate, we can see here that the boys and girls differed. This could be because they are genetically different; a nature explanation. Boys might be biologically predisposed to be aggressive, so more likely to copy aggressive models. Alternatively, boys might be more likely to copy aggressive models because they have been rewarded for aggressive behaviours more than girls have. This would be a nurture argument.

In addition to the observations, records of the remarks about the aggressive models revealed differences, both between reactions to the actions of the male and female models and between boys and girls. Some comments appeared to be based on previous knowledge of **sex-typed behaviour**, such as 'Who is that lady? That's not the way for a lady to behave. Ladies are supposed to act like ladies . . .' and 'You should have seen what that girl did in there. She was just acting like a man. I never saw a girl act like that before. She was punching and fighting but no swearing.' Whereas comments about the female model's behaviour were disapproving, those about the male model were not. This was more likely to be seen as appropriate and approved by both boys and girls, for example in comments such as 'Al's a good socker, he beat up Bobo. I want to sock like Al' and 'That man is a strong fighter, he punched and punched and he could hit Bobo right down to the floor and if Bobo got up he said, 'Punch your nose.' He's a good fighter like Daddy.'

KEY TERM

sex-typed behaviour: actions that are typically performed by one particular gender and are seen in society as more appropriate for that gender. For example, aggression is seen as a masculine-type behaviour and was more commonly imitated by boys.

Conclusions

The results strongly suggest that observation and imitation can account for the learning of specific acts without reinforcement of either the model or observer. All four hypotheses were supported:

- Observed aggressive behaviours are imitated: children who see aggressive models are likely to be more aggressive than those seeing a non-aggressive model or no model.
- Observed non-aggressive behaviours are imitated: children seeing non-aggressive models will be less aggressive than those seeing no model.
- Children are more likely to copy a same-sex model, although this may depend on the extent to which this behaviour is sex-typed.
- Boys are more likely to copy aggression than girls.

Strengths and weaknesses

The main method was a laboratory experiment. This means that it was possible to control **extraneous variables** such as ensuring there was a possibility that the children in any condition would show aggressive behaviour. This was done by showing them nice toys but then taking them to another room. Also, all children in both experimental groups saw a model for the same length of time, and in each condition their behaviours were standardised. This means the research was more **valid** – the researchers could be sure that the differences in results between conditions were due to the differences between the models – and more **reliable**, because each child within a condition experienced exactly the same exposure. **Inter-observer reliability** was also checked for both the initial observations of aggressiveness and for the data recording – and was very high. The pre-testing of the children's aggressiveness was another factor that increased **validity**, because it ensured that differences between conditions were due to the models and not to individual differences between the children who happened to be in each group.

Reflections: The procedure was standardised in many ways. Identify as many of these as you can. Justify why each one of these was necessary – try to think about what might have happened if each participant had not been treated in the same way.

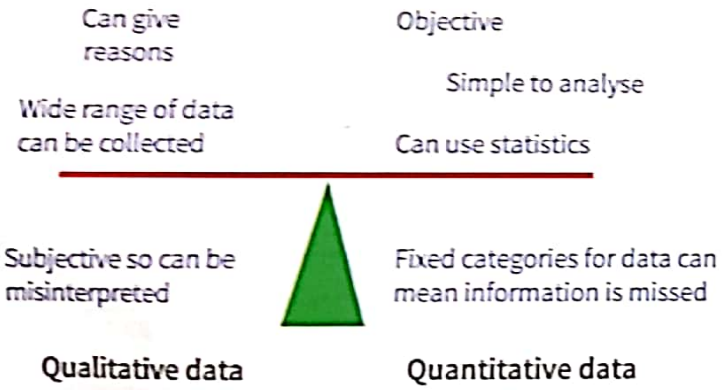
The main measure of the DV was through observation. As the observers were behind a one-way mirror, the children were unaware that they were being watched. This increases validity as they were likely to behave naturally rather than responding to **demand characteristics** as they might have done had they known they were being observed. The observation period was divided into time intervals (of five seconds) and the categories were clearly defined (e.g. imitative and non-imitative behaviours), which also helped to improve both validity and reliability.

Among the weaknesses of the study is that only six children were used in each experimental condition and, although

they were matched to reduce the risk of **participant variables** confounding the results, it is still a small sample. Furthermore, it is possible that the children were quite similar, as they all attended the same nursery based at a university, suggesting that they all had academically able parents. This could bias the **sample**, lowering validity.

The study collected both quantitative data, which was an **objective** record of the number of imitative actions in each category and qualitative data, which although more **subjective**, provided some explanation of the reasons behind the tendency of the children to copy some behaviours and not others. It could have been useful to have asked the children for self-reports of their emotions when they were observing the model or reacting towards the Bobo doll. This may have helped to further explain differences such as the influence of sex-typing on imitation. It would also have been useful to have followed the children up to see how long the children's acquired behaviours lasted. If imitation leads to learning, the change in behaviour should be relatively permanent. However, if this were so, it would also raise **ethical issues** with the study as some of the children might have been harmed by becoming more aggressive. Even if this were not the case, the children had been mildly annoyed, which could be psychologically distressing.

Reflections: Think about qualitative and quantitative data as the two ends of a see-saw (Figure 4.3). What are the advantages and disadvantages that weigh the see-saw in each direction?



4.3 The strengths of qualitative data are the opposite of those of quantitative data and vice versa, so they weigh against each other

SELF-ASSESSMENT QUESTIONS

1 The children in Bandura et al.'s study were divided into many different groups according to the various levels of the three different IVs. Draw a table to show how many children were in each of the different conditions.

2 The evidence illustrates several examples of sex-typed behaviours in the children. Identify at least **two** examples.

4.2 Core study 2: Saavedra and Silverman (button phobia)

Saavedra, L. M., & Silverman, W. K. (2002). Case Study: Disgust and a specific phobia of buttons. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(11), 1376–1379.

Aim

The aim was to examine the role of **classical conditioning** in relation to fear and avoidance of a particular stimulus. In the context of a specific **phobia**, researchers wanted to see if using a type of exposure therapy could reduce the disgust and distress associated with buttons.

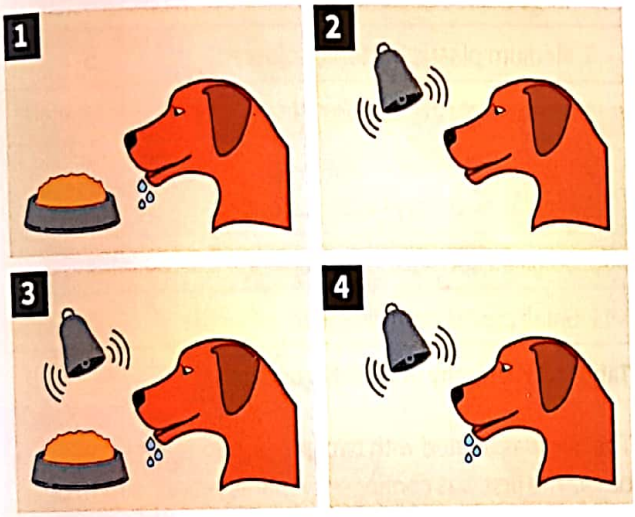
KEY TERMS

classical conditioning: learning through association, studied in both humans and animals. It is a learning process in which a new stimulus, which initially has no effect (the neutral stimulus, NS), becomes associated with another stimulus (the unconditioned stimulus, UCS). The UCS already produces a response (the unconditioned response, UCR), which is often an innate (instinctive) reaction. Following pairing of the UCS and NS, sometimes only once by more often repeatedly, the NS will produce a response similar to the existing UCR, so the NS becomes known as the conditioned stimulus (CS) and the newly learned response, the conditioned response (CR). A famous example of this process can be seen in Figure 4.4.

phobia: the irrational, persistent fear of an object or event which poses little real danger but creates anxiety and avoidance in the sufferer.

Background

Classical conditioning is a form of learning in which an unconditioned response becomes linked to a previously neutral stimulus to create a learned association. It was investigated by Ivan Pavlov, who observed dogs salivating in a laboratory as part of a totally different experiment. Salivation is an unconditioned (that is, uncontrolled) response to being presented with food (unconditioned stimulus). The dogs came to anticipate the food when they saw researchers in the lab, before feeding times. Pavlov presented the food alongside a range of neutral stimuli, such as bells, which created a learned association. Over the course of several trials, when the dogs heard the bell (conditioned stimulus) they began to salivate (conditioned response), see Figure 4.4.



4.4 Pavlov's dog and bell

1 Before conditioning: unconditioned stimulus (food) → unconditioned response (salivation)

2 Before conditioning: neutral stimulus (bell ringing) → no response

3 During conditioning: unconditioned stimulus (food) + neutral stimulus (bell ringing) → unconditioned response (salivation)

4 After conditioning: conditioned stimulus (bell ringing) → conditioned response (salivation).

Some psychologists believe that abnormal behaviour such as phobias can also be both learned (and unlearned) in the same way as any other type of behaviour. There are several subtypes of classical conditioning, including expectancy learning, in which a previously neutral or non-threatening object or event becomes associated with a potentially threatening outcome. The consequence is that the individual *expects* the negative outcome, so experiences fear in the presence of the previously non-threatening situation. Phobias may be caused by evaluative learning, that is, a kind of classical conditioning in which the individual forms an association between a previously neutral stimulus and a negative emotion, but this is a negative *evaluation*, for example disgust, rather than fear. This suggests that disgust, as well as fear, may be an appropriate target emotion for the treatment of phobias, as well as the primary emotion of fear. In the case of phobias, an emotional response such as fear or anxiety becomes associated with a particular neutral stimulus, such as buttons. A strictly behaviourist approach to classical conditioning would suggest that there is little role for thinking and feeling in this learning process. However, the case study we will look at now involves a deeper exploration of the role of cognition and emotion in changing phobic behaviour, known as **evaluative learning**.

KEY TERM

evaluative learning: a form of classical conditioning wherein attitudes towards stimuli are considered to be the product of complex thought processes and emotions which lead an individual to perceive or evaluate a previously neutral stimulus negatively. Attitudes acquired through evaluative learning may be harder to change than more superficial associations.

From their study of adults with a blood phobia, Hepburn and Page (1999) suggested that treating patients' disgust, as well as their fear, would helped them to make progress. De Jong et al. (1997) worked with children who had a spider phobia. Although no attempt was made to manipulate feelings of disgust, their feelings of disgust declined alongside their reduction in fear.

ISSUES AND DEBATES

nature versus nurture – in classical conditioning, the environment is considered the main influence on behaviour. Individuals are considered to be 'blank slates' when they are born, and are shaped purely by nurture.

Method

Research method and design

This was a clinical **case study** as it involved just **one participant** whose life history and treatment was studied in depth. Data was collected using self-report measures. Both the boy and his mother were interviewed by the researchers about the onset of his phobia and his

subsequent behaviour. The results of the treatment were measured using a nine-point scale of disgust known as the 'Feelings Thermometer'.

RESEARCH METHODS

Case studies often collect a large amount of qualitative data; this is often the main reason they are chosen as a method. In this study, interviews produced a history of the boy's experience with buttons and helped researchers understand the origin of his phobia.

Sample

The participant was a 9-year-old Hispanic American boy. Along with his mother, he had sought support from the Child Anxiety and Phobia Program at Florida International University, Miami. He met the criteria for having a specific phobia of buttons and had been experiencing symptoms for around four years prior to the start of the study.

Procedure

The boy and his mother both provided informed consent to participate in the study. They were interviewed in order to determine whether any trauma or abuse could explain the boy's phobia. The phobia had begun at age five, when the boy had knocked over a bowl of buttons in front of his class and teacher. He found the incident distressing, and from that time onwards his aversion to buttons steadily increased. When he was interviewed the phobia was interfering significantly with his normal functioning; he could no longer dress himself and had become preoccupied with avoiding touching buttons or clothing that could have touched buttons.

RESEARCH METHODS

This research method involves studying **one or a very small number of participants** (such as a family unit) in great depth. It is particularly suitable in researching phenomena that are unusual or rare.

It was necessary for the researchers to understand the boy's specific feelings towards buttons prior to starting treatment. Through discussion with the participant, they created a hierarchy of feared stimuli, with each item on the list provoking increasing fear (see Table 4.2). The most difficult items for the child were small, clear plastic buttons. These were rated at an '8' on the nine-point Feelings Thermometer. Handling these or touching someone wearing them was the most unpleasant task for the boy.

Stimuli	Distress rating (0–8)
1. Large denim jean buttons	2
2. Small denim jean buttons	3
3. Clip-on denim jean buttons	3
4. Large plastic buttons (coloured)	4
5. Large plastic buttons (clear)	4
6. Hugging Mom when she wears large plastic buttons	5
7. Medium plastic buttons (coloured)	5
8. Medium plastic buttons (clear)	6
9. Hugging Mom when she wears regular medium plastic buttons	7
10. Small plastic buttons (coloured)	8
11. Small plastic buttons (clear)	8

Table 4.2 Hierarchy of fear/disgust

The boy was treated with two interventions, one after the other. The first was contingency management, a form of **positive reinforcement** therapy. This was a behaviour-focused approach which meant the boy was rewarded for showing less fear and for actually handling the buttons. The positive reinforcement was given to the boy by his mother only after he had completed a gradual exposure to buttons. These treatment sessions lasted between 20 and 30 minutes.

KEY TERM

positive reinforcement: a form of operant conditioning, another theory of learning. It involves rewarding desirable behaviour to encourage it to be repeated. For example, praising a child for saying 'please' and 'thank you' positively reinforces good manners.

Reflections: Imagine a different phobia, one of spiders, mice or peanut butter, for example. Suggest eight different scenarios that could be imagined by a person receiving similar treatment to that described here.

The second form of therapy, and the main focus of the study, was known as 'imagery exposure'. Interviews with the boy had revealed that he found buttons touching his body disgusting, and he also believed that buttons smelled unpleasant. These ideas formed the basis for disgust imagery exercises. Unlike *in vivo* exposure, where the individual actually physically handles or is exposed to fearful stimuli, imagery exposure therapy uses visualisation techniques.

Disgust-related imagery exposures were incorporated with cognitive **self-control** strategies. The boy was asked to imagine buttons falling on him, and to consider how they looked, felt and smelled. He was also asked to talk about how these imagery exposures made him feel. The exposures progressed from images of larger to smaller buttons, in line with the boy's fear hierarchy.

KEY TERM

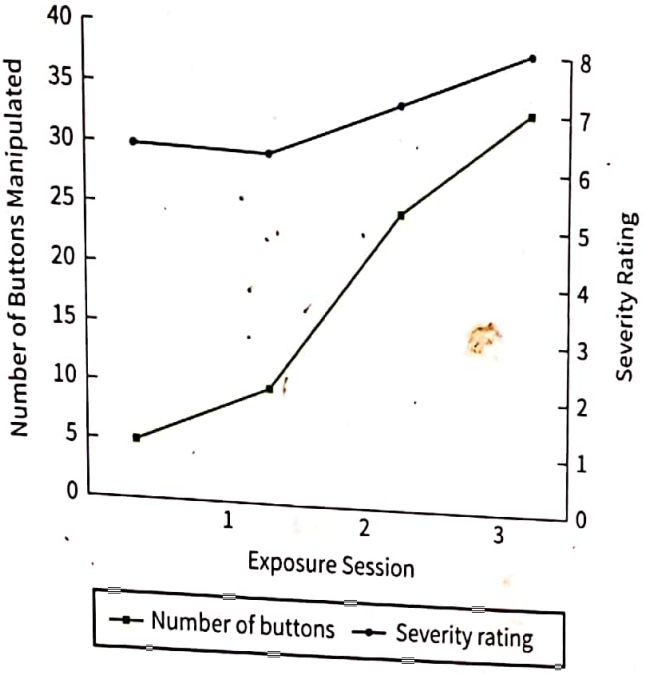
self-control: a form of cognitive behavioural therapy. It involves using 'self-talk'; the individual is taught to recognise difficult situations, acknowledge troubling thoughts and consider alternative, positive thoughts.

Results

Positive reinforcement therapy

The outcome of this therapy was a successful completion of all the exposure tasks listed in the hierarchy of fear. The boy was also observed approaching the buttons more positively. One example of this was that he started handling larger numbers of buttons during later sessions.

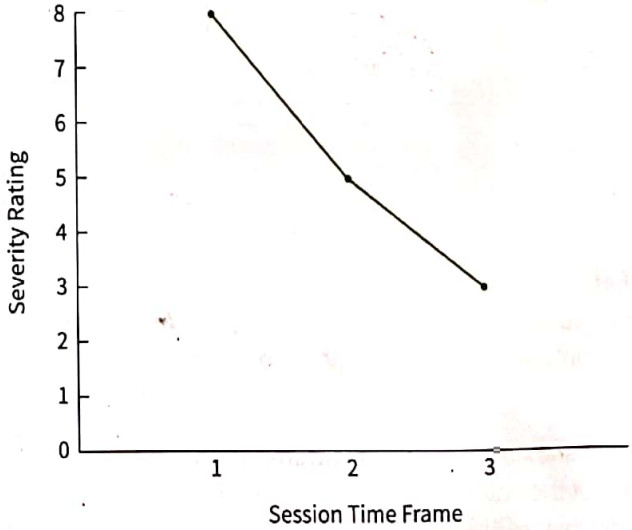
However, his subjective ratings of distress increased significantly between sessions two and three, and continued to rise (see Figure 4.5). By session four, a number of items on the hierarchy such as hugging his mother while wearing buttons had increased in dislike from the original scores. So despite his behaviour towards the fearful stimuli improving, his feelings of disgust, fear and anxiety actually increased as a result of the positive reinforcement therapy. This finding is consistent with evaluative learning; despite apparent behavioural change, evaluative reactions (i.e. disgust) remain unchanged or even increased.



4.5 Ratings of distress: positive reinforcement therapy

Imagery exposure therapy

This appeared to be successful in reducing the boy's rating of distress. One example of this is shown in Figure 4.6, relating to imagery of 'hundreds of buttons falling all over his body'. Prior to imagery therapy, the boy rated this experience the most fearful and disgusting (score of 8 on the Feelings Thermometer). This reduced to 5 midway through the exposure, and just 3 after the exposure was complete.



4.6 Ratings of distress: imagery exposure therapy: (1) before, (2) midway and (3) after imagery exposure

Following his treatment, six month and 12 month follow-ups were conducted. At these assessment sessions, the boy reported feeling minimal distress about buttons. He also no longer met the diagnostic criteria for a specific phobia of buttons. His feelings towards buttons no longer affected his normal functioning; he was also able to wear small, clear plastic buttons on his school uniform on a daily basis.

Conclusions

The researchers concluded that the treatment was successful. In particular they argue that:

- emotions and cognitions relating to disgust are important when learning new responses to phobic stimuli
- imagery exposure can have a long-term effect on reducing the distress associated with specific phobias as it tackles negative evaluations.

Strengths and weaknesses

This piece of research involved a case study. This means that the sample is small (in this case, one person) and difficult to generalise from. As the participant was diagnosed with a specific phobia of buttons, it makes the case even less likely to be representative of the general population. However, the case study is highly **valid**; the participant was studied over a period of time using several different methods of data collection. The researchers used **standardised** measures such as the Feelings Thermometer before, during and after therapy. Collecting and analysing **quantitative data** means we can be quite sure that the improvement seen in the phobic reactions of the little boy was highly likely to be a result of the treatment he received.

Reflections: This study used only one participant. Can you explain why the results are still useful, despite possibly not being representative of others?

A substantial amount of **qualitative data** was also gathered about the boy. An example of this was the background information obtained by interviewing the boy and his mother about the button incident at his school. This type of data is useful because it can help us to understand the reasons underlying abnormal behaviour.

In many ways, this study could be considered to be **subjective**. The participant created his own hierarchy of fear and disgust relating to his button phobia. He also gave personal ratings which were highly individual to his own thoughts and feelings. However, as the aim of this study was to understand the experience of evaluative learning in an individual with specific phobia, these measures were appropriate. A scale created by the researchers for use with all phobic patients would not have been as relevant or informative about personal progress.

Working on a case study involves building rapport with the participant. There is less room for anonymity or objectivity. This means there is a higher risk of **bias** which may compromise the **validity** of the study. There may be researcher bias as the researchers select and report on a particular participant (who may be more likely to have a positive outcome). **Demand characteristics** may be more obvious also. For example, the boy was fully aware he was undergoing therapy with the intention of improving his phobic symptoms. This might have affected the ratings he gave to the different levels of exposure therapy.

When using children as participants, **ethical issues** can be a major concern. In this instance, the boy and his mother gave informed consent to participate in the study. This was important as the therapy involved deliberate exposure to distressing stimuli, whether real or imagined. Overall, the aim of the study was to improve the boy's quality of life and minimise psychological distress, which is less concerning. In addition, the boy's anonymity was preserved, which allowed him to maintain his privacy.

SELF-ASSESSMENT QUESTIONS

3 Briefly explain why the study by Saavedra and Silverman was carried out.

4 Describe the conclusions from this study.

4.3 Core study 3: Pepperberg (parrot learning)

Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African Grey parrot (*Psittacus erithacus*): Learning with respect to categories of color, shape, and material. *Animal Learning & Behavior*, *15*(4), 423–432.

Aim

The aim of this study was to see whether a parrot could use vocal labels to demonstrate a symbolic understanding of the concepts 'same' and 'different'.

Background

Humans are not the only animals capable of making meaningful communication. There has long been interest in non-human primates, such as chimpanzees and gorillas, and their abilities to talk to one another. Researchers have even been able to teach them to express their ideas and needs through pointing at symbols or using sign language (see Figure 4.7). Some of these studies have produced surprising results; namely that non-human primates have the cognitive capacity to express abstract ideas and, in some cases, form meaningful sentences.

4.7 Teaching sign language to a primate

Although some language skills appear to be confined to primates, one specific cognitive capacity is thought to be present in other species. The conceptual categorisation of items as the same or different is thought to be present in other animals. Identifying items as the same or different requires a particular set of cognitive abilities. Firstly, the animal must recognise the category that is being shared (for example, that both items are green in colour). From this single attribute, they must also realise that this sameness can be applied to other items that they have not yet encountered. This requires a skill known as symbolic representation. This means the concept of sameness or difference learned from one experience can be applied to an entirely new and different situation.

The focus of much language training has been on non-human primates because their advanced cognition makes them useful participants in such research. However, there is evidence to suggest that the capacity to recognise same/different items is present in many bird species. This may be the case as it serves an adaptive function; it benefits a bird's survival.

Method

Research method and design

This was an animal case study involving one subject who was trained and tested over a couple of years.

Sample

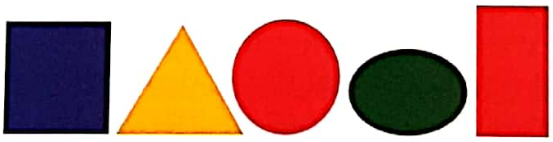
This study focused on one African Grey parrot (Figure 4.8) called Alex. Alex had been involved in prior research on communication and cognition for around ten years. During the day, the parrot had free access to all areas of the laboratory and at night he was confined to a wire cage. He was fed a diet suitable for his species and given toys to play with.

4.8 African Grey parrot

Procedure

Prior to training, Alex already had a considerable vocabulary as a result of his previous experience as a research subject. For example, he could already name the colours red, green, yellow, blue and grey, several shapes (e.g. triangle, square) and different kinds of material (wood, cork, hide, paper and some metals). He also had experience of replying to verbal prompts. When asked 'what colour?' he could often correctly name the colour of an item presented to him. He could even combine responses to describe items, e.g. to name 'green wood'.

The purpose of training was to teach Alex to respond to questions with a categorical label, rather than simply describing an item's appearance, or stating whether it matched or did not match a paired item. This is a more complex task which required abstract thinking. He engaged in training sessions two to four times a week, each session lasting between five minutes and one hour. Alex also engaged in training for other studies during this period.



4.9 Different shapes were used in the trials

The training method used in this study is known as the model/rival or M/R approach. It is based on the concept of modelling demonstrated in the first study in this chapter by Bandura et al. (1961). One human acts as the trainer of the second human by presenting the second with objects (Figure 4.9), then asking questions about the objects and offering reward or praise to desirable responses. Thus, the second or learner human acts as a model to the parrot who is watching the interaction. The parrot can be considered a 'rival' for the trainer's attention; if they offer the right vocalisation when the question is asked, they receive the reward or praise rather than the model human. The roles of trainer/model are then reversed.

At the beginning of training, a system of **continuous reinforcement** was used. This was intended to create the closest possible association between the object or category and label to be learned. During training on same/different, the trainer would ask the model 'what's the same?' or 'what's different'? The model would either respond with the correct category label and be rewarded by being given the object, or would give an incorrect response and be scolded and have the object taken away.

KEY TERM

continuous reinforcement: when a learner receives a reward each time they perform a desirable behaviour. It is one of several possible schedules of reinforcement.

RESEARCH METHODS

It was important in terms of **validity** that Alex's responses could be understood. This is why testing was delayed. If testing had been started earlier, and any sounds Alex made that were a bit like 'matter' were counted as correct, this may have resulted in an over-estimate of his ability.

Task

At the start of the study, Alex could already say the labels 'colour' and 'shape'. However, it took some time to master the vocalisation for matter (he pronounced 'mah-mah') so the testing phase was delayed for the purposes of accuracy. Alex was tested by secondary trainers who had not worked with him on learning same/different. The materials were paired from a selection given to a student who had nothing to do with the study, in order to create an unbiased set of stimuli. They also randomly ordered the set of questions used in each trial. Although researchers were only interested in data from the same/different questions, other questions were included to prevent boredom effects such as:

- What colour?
- How many?
- What shape?

In each trial, Alex was presented with two objects that could differ in one of three categories: shape, colour or material. For example, the group might include a blue wooden triangle and a blue wooden square. In some trials, Alex was asked by the trainer: 'what's the same?' and 'what's different?' The correct response would be for Alex to name the categories that were the same (in this example colour and material), and those which were different (shape). The task would either involve pairs of familiar items (familiar trial) or one or both items which he had not yet encountered in training (novel trial).

Reflections: Why do you think Alex was asked to make judgements on items that had three dimensions (shape, colour and material) rather than just two?

The principal trainer was present in each trial, but sat facing away from Alex and was unable to see the objects being presented. After each of Alex's responses, the trainer repeated the vocalisation aloud. If Alex's response had been correct, he was rewarded with praise and given the items. Getting the answer right first time counted towards the 'first-trial' response rate. If it was an incorrect or indistinct vocalisation, Alex was told 'No!' and had the object removed while the trainer turned their head away from the parrot (known as a 'time-out'). This correction procedure was repeated until the correct response was given, and the number of errors was recorded.

Reflections: Alex's primary trainer was present in each trial. However, they didn't look at the parrot and were asked to repeat his responses. Can you explain why they were given this role?

Results

In tests involving familiar objects, Alex correctly responded to 99/129 trials (76.6%). This was for first trials answered correctly and those that involved correction procedures. For first trials only, Alex answered correctly in 69/99 instances (69.7%). However, this finding gives a conservative picture of the success of the training. One reason for this is that the response was only counted as correct if it was the first vocalisation Alex gave in response to the question; but rather than giving the 'wrong' answer on the remaining trials, Alex might not respond with an answer at all, but instead make a request for another object or actions unrelated to the trial.

RESEARCH METHODS

In recording Alex's responses, an **operational definition** was used. The response was only counted as correct if it was the first vocalisation Alex gave. Operationalisation of variables helps to improve reliability because it means that the experimenters are likely to be more consistent in the way that they collect data.

The tests also involved novel objects which measured the extent to which Alex could generalise his understanding and communication around the concepts of 'same' and 'different' to entirely new situations. Here Alex actually performed slightly better on the task, scoring 96/113 on all trials (85%), and 79/96 (82.3%) on first-trial performance only. While we might expect Alex to have found it more difficult to judge items which he had never encountered before, he was actually more accurate at doing so. One reason the researchers suggest for this difference is that Alex received the items involved in each trial as his reward. As such, he might have been motivated to do better on novel trials because he was curious to investigate newer reward items.

Conclusions

The researchers concluded from the case study of Alex that after training:

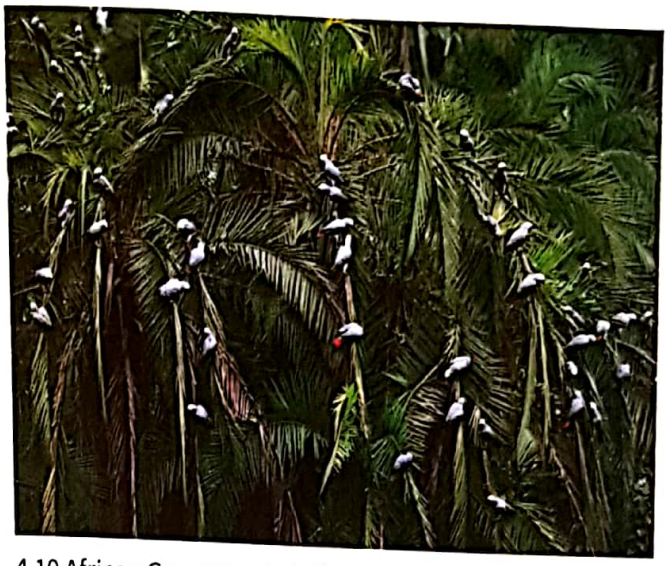
- parrots have the potential to demonstrate comprehension of the symbolic concepts 'same' and 'different'
- they may also learn to respond to verbal questions to vocalise categorical labels.

Strengths and weaknesses

This study involved one male African Grey parrot, making it a case study. Like the previous piece of research on phobias, this method makes it difficult to generalise from. As the parrot was a trained laboratory animal, it would be hard to say that it is representative of the general population. The researchers mention that Alex suffered from boredom at times, hence they varied his training and testing schedules in order to **control** for repetitive behaviours that were unrelated to the leaning process. This species is intelligent and prone to self-injurious behaviour in captivity as a result of boredom; a major challenge for those conducting research with them.

Reflections: How do you think the lab-based environment differs from the parrot's natural environment (Figure 4.10)? What effect might this have on their behaviour?

There were attempts to maintain the **validity** of the research; for example, the question order and materials were selected by a student with no connection to the project. Similarly, the trainer conducting the trials had not trained Alex on the same/different task. Both of these controls meant that researcher bias was limited.



4.10 African Grey parrots in the wild

It also meant Alex did not respond to any **demand characteristics** from interacting with his usual trainer. The research involved collecting **quantitative data** in the form of correct responses to the same/different questions. This allows us to make an objective analysis of whether or not Alex was able to understand the abstract concepts. Usefully, it also enabled comparison of two different sorts of tests, one involving familiar objects and one involving novel objects, so researchers could establish whether the rules of same/different had been generalised beyond training materials.

When using animals as participants, different sorts of **ethical issues** must be considered. In this instance, Alex is reported to be well treated and does not appear to have been physically harmed as a result of the research. There is no suggestion he was underfed or understimulated in order to encourage his participation in training. For example, Alex's conditions are described in a way that suggested he had adequate space to explore the laboratory for a considerable part of the day, and was given toys and interactions to prevent boredom. However, the researchers acknowledge that the species of parrot involved in the study is fairly intelligent, and being kept in an artificial environment for such an extended period of time while participating in non-naturalistic behaviours could be considered unethical.

SELF-ASSESSMENT QUESTIONS

5 Outline the aim of the study by Pepperberg.
6 Discuss **two** strengths and **two** weaknesses of this study.

4.4 Issues, debates and approaches

The application of psychology to everyday life

There are clear implications from Bandura et al.'s work. Children all over the world are exposed to aggression, both real, such as domestic violence in the home or violence on the news, or fictional, such as in cartoons, films and computer games. This study shows how such models can influence the behaviour of children, especially of boys. One consequence of the recognition of such potential effects is that many countries have restrictions on viewing, such as TV times before which programmes with inappropriate content for children cannot be shown (this is called the 9 p.m. 'watershed' in the UK) and certifications for films and games which indicate the age of child it is deemed suitable for viewing.

Similarly, the study by Saavedra and Silverman shows how therapy based on the principles of classical conditioning can be used to treat specific phobias. A phobia is a distressing mental health condition which can negatively affect people's quality of life. Methods such as disgust imagery exposure are used in clinical practice to challenge the fearful associations with phobic stimuli. This piece of research demonstrates the potential long-term improvement that can result from exposure therapies.

The research by Pepperberg into parrot cognition and communication raises interesting questions. It shows how non-primate animals can also be taught to communicate using modelling and reinforcement. Using these learning techniques has enabled us to see that other animals are able to reason about abstract categories, and generalise such concepts to novel situations.

Nature versus nurture

In the study by Bandura, situational influence of models on the acquisition of aggressive behaviours (and the suppression of them by the non-aggressive model) are examples of nurture. However, the differences between the responses of boys and girls to the same models could be explained by either nature or nurture. Boys were more likely to imitate aggressive behaviour than girls, possibly because boys have more of the hormone testosterone which is a nature factor. Alternatively (or additionally), boys may be more likely to imitate aggression because they have already acquired stereotypes about what is acceptable behaviour for males or have been rewarded for masculine-type behaviours. These would be examples of nurture.

The theory underlying the acquisition and treatment of phobias in the study by Silverman and Saavedra is classical conditioning. Classical conditioning relies solely on a nurture-based explanation of learning. Phobias are not considered innate or genetically inherited. Instead they are considered to be products of negative experiences with previously neutral stimuli. Treatment is based on the same principles; that subsequent neutral or positive experiences with the phobic stimuli (along with cognitive therapy) can reduce fearful responses.

Once again, both the theories of operant conditioning and social learning demonstrated in Pepperberg rely on a nurture-based approach to learning. As human training shaped the parrot's behaviour through rewards, it was able to demonstrate an important aspect of its cognition. However, there are clearly differences in the parrot's abilities and those of humans and other primates which cannot be attributed to factors in the environment, but rather in-born traits and abilities unique to different species.

Individual and situational explanations

This debate is relevant to the study by Bandura. Imitation clearly suggests that situational factors matter in that the model is an aspect of the situation, as are differences between male and female models. However, individual factors could also explain some differences in imitation. Individual factors in operant conditioning can explain why, even when girls and boys are exposed to the same models, their acquisition of behaviours differs because boys and girls may be differently rewarded for sex-typed behaviours. For example, a daughter may be praised for *not* fighting but a son praised for 'sticking up for himself'.

Individual and situational explanations can also be considered in the context of the study by Pepperberg. Alex the parrot's abilities in the familiar and novel testing did show significant differences; namely a higher rate of success on object pairs which were unfamiliar to Alex in some way. This suggests that situational factors (the familiarity of objects) can affect an animal's ability to offer a correct response.

The use of children in psychological research

The children used in Bandura et al.'s study did not appear to have been given the opportunity to consent to the study, or to withdraw. Since children are particularly vulnerable, and this study had the potential to cause distress, if this were the case these would be issues for ethical concern. Although the headteacher at the nursery school is thanked in the study, so she was clearly aware of the procedure,

there is no indication of whether the parents' consent was obtained. When children are used in studies, ethical guidelines typically suggest that parents' or guardians' consent should be obtained in addition to the child's own.

On a practical level, the use of children rather than adults in Bandura's study was ideal. Children have been exposed to much less violence than adults and there are likely to be fewer extraneous factors affecting their aggression levels (such as a bad day at work). In general, children are more naive than adults, so the participants would have been less likely to suspect that they were being shown aggressive models in order to investigate the effects of these on their own behaviour. These considerations all lead to the greater potential for representative effects of the procedure on children than if the same study were conducted with adult participants.

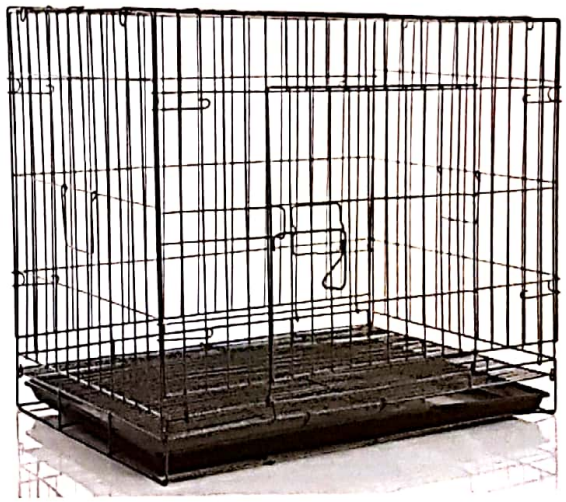
In the study by Silverman and Saavedra, the child participant (aged nine) was asked for consent. In accordance with ethical guidelines, his mother also consented to his participation. This study was potentially highly distressing, as it involved both real and imagined exposure to frightening stimuli. Furthermore, the boy could be considered vulnerable as his specific phobia is a recognised mental health condition. However, the intention of the researchers was to treat his phobia and improve his quality of life, which may justify the temporary distress caused during treatment.

The use of animals in psychological research

Although the Bandura study was conducted on humans, similar research could, and has, been done with animals. Clearly there are some reasons why this would be better – it removes the potential risk of causing short-term distress or long-term harm to children. In addition, although it is assumed that the children remained naive to the purpose of the activity, it is still possible that they believed that they were 'supposed' to copy the adult's behaviour. This would not be a risk if animals were used, and the possibility that some animals had been exposed to more aggression prior to the study could be controlled.

Conversely, there would be disadvantages. Although there are biological sex differences in the behaviours of animals, they do not acquire sex-typed behaviours that are a consequence of cultural factors in the way that children do. Because this has been shown to be important, an animal study would not be able to test such social influences.

The research by Pepperberg was a case study of the training and testing of a single male African Grey parrot. Use of this participant was necessary as the researchers were particularly interested in the cognitive abilities of non-primates. Practically there were some advantages to working with this animal; it was relatively easy and cheap to care for. It had already been involved in training for other studies and was known to respond well to training in vocabulary exercises. Furthermore, there is no evidence that Alex was harmed as a result of the learning techniques of modelling and positive reinforcement. The study was conducted over a number of years and involved training on at least a weekly basis, which would have been difficult to achieve with a human participant. The parrot was also motivated by receiving the reward of the object in each trial; again motivating a human participant to continue with on-going trials might have been harder.



4.11 Is it right to keep animals in captivity for the purposes of research?

On the other hand, there are some practical limitations to working with animals in studies of this kind. The researchers had to devise a method for testing the cognitive abilities of the animal without having direct access to that animal's understanding. The parrot was unable to explain its reasoning in great depth. Fortunately the design of the task meant the investigation went beyond a simple match or no-match test, and the parrot was also able to demonstrate its learning in novel situations. There are also ethical concerns around keeping an animal in captivity away from its natural environment (Figure 4.11), and engaging it in artificial tasks such as imitating human speech.

Summary

Bandura et al.'s study used adults being aggressive to a Bobo doll to show that children's behaviour can be affected by that of a model. Exact aggressive behaviours were imitated although non-aggressive modelling was also effective. Children were more likely to copy a same-sex model and boys engaged in more aggressive imitation than girls. This was a well-controlled laboratory experiment measuring the dependent variable through objective observations which were reliable. Qualitative data suggested that the children recognised sex-typing and were surprised by behaviour that did not fit the pattern. The findings suggest practical applications for protecting children, e.g. through film certification.

The study by **Saavedra and Silverman** aimed to investigate the role of evaluative learning in treating a specific phobia of buttons. They used a case study of a young boy to explore the origin of his phobia, and treated him using both reinforcement and imagery exposure therapies. He responded best to the imagery exposure technique which relates closely to the thoughts and feelings associated with learned responses. This was a unique piece of research which gained both qualitative and quantitative data about the participant's progress through therapy. It has practical applications for the treatment of specific phobias, e.g. through imagery exposure relating to disgust.

Pepperberg's study used a single male African Grey parrot to show that birds of this species are capable of learning through modelling and reinforcement. Through a lengthy process of training and testing, researchers were able to establish that the animal could comprehend and apply the cognitive constructs of 'same' and 'different' to both familiar and unfamiliar objects. This was a lab-based study which had limited generalisability and lacked ecological validity. Using objective, quantitative measures of correct responses however did show that abstract categorical constructs may not be limited to humans and other non-human primates.

Exam-style questions

1 In their conclusion, Bandura et al. suggest that social imitation can speed up the learning of new behaviours as they can be acquired without the need for reinforcement, that is, without operant conditioning. Suggest why acquiring new behaviours through social imitation would be quicker than through operant conditioning. **[2 marks]**

2 If Bandura et al.'s study were performed today, the researchers would be required to obtain informed consent from both the children and their parents or guardians. Explain why this would be necessary. **[3 marks]**

3 There are several explanations for how learning occurs, including classical conditioning, operant conditioning and social (observational) learning. In the study by Saavedra and Silverman on button phobia, one technique for therapy used operant conditioning.

a Explain what is meant by operant conditioning. **[3 marks]**

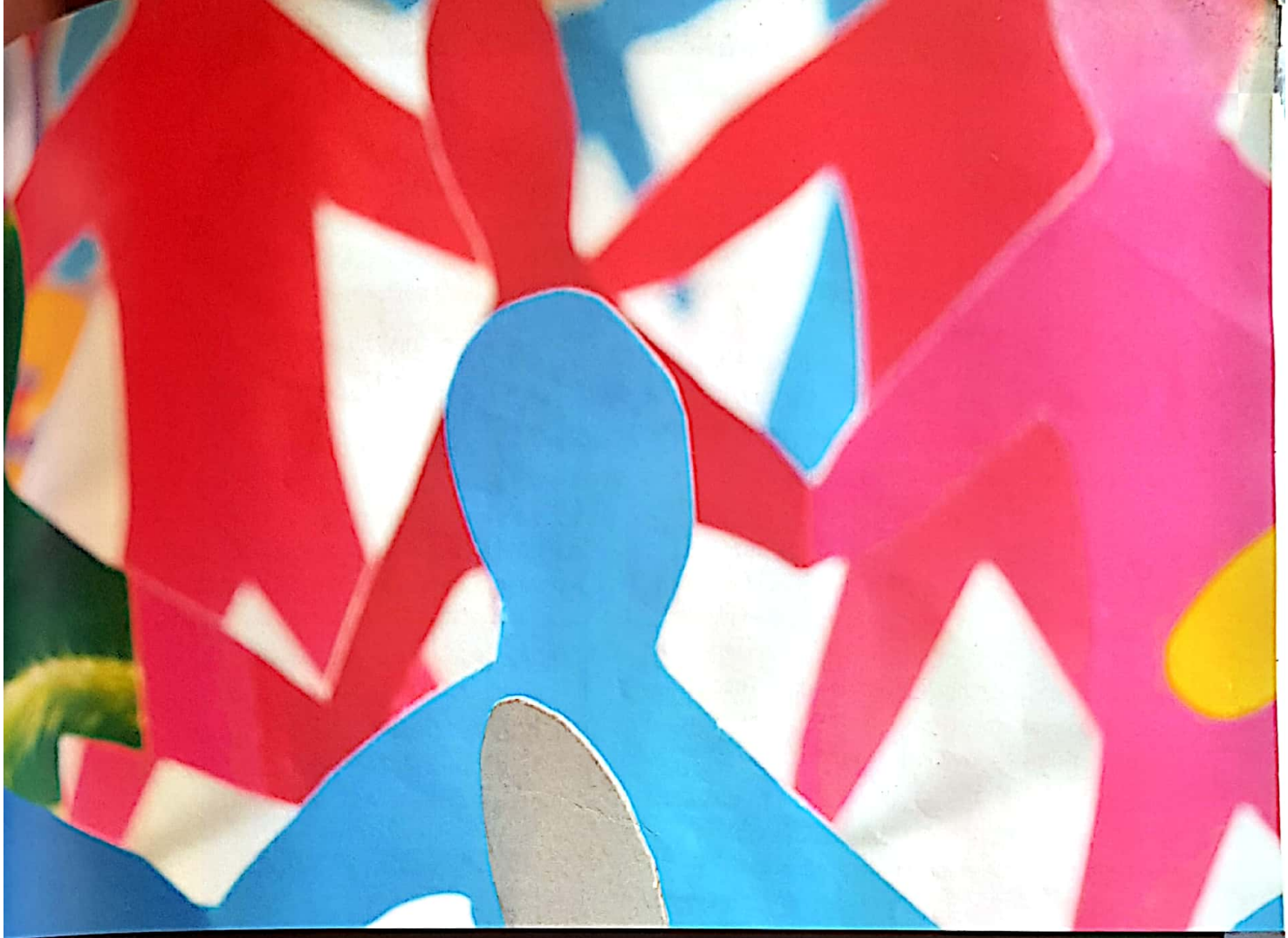
b Describe how operant conditioning was used to help reduce the boy's phobia in the study. **[2 marks]**

4 One debate in psychology is the nature versus nurture debate. This is illustrated in the study by Pepperberg (parrot language).

a Explain what is meant by 'nature' and 'nurture'. **[2 marks]**

b One cognitive capacity present in many animals is the ability to understand that things can be 'the same' or different'. This categorisation requires the symbolic representation of 'sameness'. This ability is found in many species. Explain why this suggests it is controlled by '**nature**'. **[2 marks]**

c Describe **one** way that '**nurture**' is illustrated in Pepperberg's study. **[4 marks]**



Chapter 5
Social approach

Introduction

The aim of this chapter is to introduce you to the social approach to psychology and to explore three studies from this approach. They are:

- **Milgram (obedience)** which is based on the conflict between individual conscience and obedience to authority and considers how far a person would obey instructions which involved hurting another person
- **Piliavin et al. (subway Samaritans)** which looks at how bystanders behave in real-life situations and factors that affect their desire to help, including diffusion of responsibility
- **Yamamoto et al. (chimpanzee helping)** which considers pro-social behaviour and instrumental helping in chimpanzees, and looks at whether chimpanzees have the ability to help others based on specific needs.

These three studies serve to illustrate the main assumptions of the social approach which are that:

- behaviour, cognitions and emotions can be influenced by other individuals
- behaviour, cognitions and emotions can be influenced by groups or social contexts.

Throughout the chapter you will find out about how the presence of others can influence the thinking and behaviour of humans and animals. In each core study you will learn about relevant background to the research, how the study was carried out, the findings and the conclusions. You will also see how research methods in psychology are used in different ways within the social approach. You will explore how a range of issues and debates can be applied to the concepts considered in this approach.

Just fitting in?

Imagine you are walking to a friend's house. You notice an older person about to cross the road into busy traffic. Perhaps you might stop and ask if they need help. Perhaps you would like to help but do not feel confident approaching them, or maybe you do not feel it is your place to help. You might even have felt bad about just walking away from the situation. Have you ever wondered why and how we make these decisions about interacting with others?

Many factors are involved when we consider how we behave in social situations. Social psychologists argue that all of our thoughts, feelings and behaviour take place in the real or imagined presence of others. Do you agree? Consider whether you are the same 'self' in private, with friends and family, at school, work or in public. The social approach to psychology is concerned with many group processes in our daily lives, including prejudice, conformity, **obedience** and helping.

Some social behaviours might seem so normal and everyday that you might wonder why psychologists are concerned with them. We are surrounded by authority figures, including parents, teachers and politicians, who make and enforce rules for us to follow. We do much of this without even noticing; following signs which remind us to 'Keep off the grass' or 'Keep out' is an act of obedience.

Throughout history, and indeed, still today we are surrounded by examples of obedience to authority which has crossed from the everyday to the extreme. This is known as '**destructive obedience**'. What exactly then are our motivations for doing something that others tell us to do, even when it goes against our personal beliefs? Can you think of some historical examples of destructive obedience?

Reflections: In one of the most shocking and infamous acts of civilian violence during the Vietnam War, a group of American soldiers carried out a mass murder in the small village of My Lai in March 1968. Read more about the cause of this massacre. What explanations might there be for this kind of destructive obedience?

KEY TERMS

obedience: following a direct order from a person or people in authority.

destructive obedience: obedience that has potential to cause psychological or physical harm or injury to another.

5.1 Core study 1: Milgram (obedience)

Milgram, S. (1963). Behavioural study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378.

Aim

The aim of the study was to investigate how obedient individuals would be to orders received from a person in authority. Specifically, Milgram wanted to know whether people would be obedient even when it would result in physical harm to another person. To test this, he arranged his laboratory-based procedure to involve administering electric shocks to a victim under the orders of a researcher.

Background

The concept of obedience, in particular destructive obedience, had a particular relevance to Milgram. In Europe around the time of the Second World War, 11 million innocent people were systematically murdered on command by the Nazis during Adolf Hitler's regime (Figure 5.1). The Holocaust took the lives of those from a range of minority groups, including six million Jews. For such a large-scale atrocity to be possible required the obedience of many, indeed a majority of ordinary citizens as well as Nazi officers and guards. Some of those later tried as war criminals in the Nuremberg Trials argued in their defence that they were 'just following orders'.

5.1 Why did so many people obey the orders to commit mass murder in Nazi concentration camps?

One theory used to explain the tragic events of the Holocaust is that German citizens possessed some defective personal traits which made such extreme levels of obedience possible. This is a **dispositional** argument which seems to suggest that Germans are somehow different from others. Stanley Milgram, who was himself born into a Jewish family, sought to challenge this hypothesis. He suggested a **situational** explanation for obedience; that many people who found themselves in a similar situation would harm or even kill other human beings under the orders of an authority figure.

Prior to his study, Milgram told psychology students and some of his own colleagues about the procedure he would use involving destructive obedience, and asked them how many participants would apply the maximum voltage shocks. Those asked believed that less than 3% of participants would deliver the maximum voltage shock, with many stating they felt that no one would deliver such strong punishment.

ISSUES AND DEBATES

The **dispositional/situational** debate considers whether the person's individual characteristics or the conditions of their environment are more influential on levels of obedience.

Method

Research method and design

This study is perhaps best described as a **controlled observation**. It took place in a laboratory setting where all the variables and measurements were controlled, while the behaviour of participants was observed and recorded. Milgram originally described his study as a laboratory experiment. In this particular study each participant underwent the same procedure and there was no control condition. However, he later replicated the procedure in other studies using different variations to allow comparisons to be made.

In this study, participants' levels of obedience were measured through observation. This was operationalised as the maximum voltage given in response to the orders. Observers also noted the participants' body language and any verbal comments or protests made throughout the procedure.

RESEARCH METHODS

A **controlled observation** is where the researcher watches and records the behaviour of participants in an artificial environment, in contrast to a naturalistic observation which takes place in the individual's normal environment.

Reflections: Milgram collected both quantitative and qualitative data in this study. Can you identify these from the description given here? Consider why both types of data might be useful in understanding destructive obedience.

Sample

A newspaper advertisement was used to recruit 40 men between the ages of 20 and 50 years old. This meant it was a **volunteer sample**, composed of those who lived in the New Haven area of the United States. The men came from a range of different backgrounds and occupations, and represented unskilled workers, **white collar workers** as well as professionals.

KEY TERMS

white collar workers: refers to individuals who work in professional occupations, as compared to 'blue collar' workers which refers to those who perform manual work.

confederate: someone who is playing a role in a piece of research and has been instructed as to how to behave by the researcher.

RESEARCH METHODS

A **volunteer sample** is also known as a 'self-selected' sample. Participants are recruited through advertisements which can be a quick and easy recruitment tool. However, they may not attract a particularly representative sample.

Procedure

After responding to the newspaper advertisement, each participant was promised $4.50 for taking part. This was not conditional on their completing the study, but simply for being willing to participate. The study took place at Yale University, in a modern laboratory. The location was

chosen in order to make the procedure seem legitimate, an important situational factor in obedience.

The participants arrived individually to the lab, and were then introduced to another man whom they believed to be another participant. This man was in fact a **stooge** or **confederate**; he was a likeable, middle-aged man who worked for Milgram and had been trained in the procedure which followed. Both men were told that they would be allocated the roles of 'teacher' or 'learner' in what was to be an experiment about the effects of punishment on learning. They drew pieces of paper from a hat to determine the roles, but it was fixed so that the real participant was always allocated the role of teacher.

Reflections: Why might Milgram have gone to the trouble of giving slips of paper to the stooge and real participants to 'allocate' the roles of teacher and learner?

Next the participant was taken to another room, where the stooge was strapped to a chair and had electrodes attached to him by the experimenter. The participant was presented with the shock generator (see Figure 5.2), which consisted of rows of switches labelled with voltage readings ranging from 15 V to 450 V. The shock voltage was also labelled in ascending order with words such as 'moderate shock', to 'danger: severe shock' and for the final two switches 'XXX' (Figure 5.3).

5.2 Milgram's participants were shown the shock generator before the procedure began

5.3 The simulated shock generator used in the study

The participant was told that although the shocks were painful, they were not dangerous. They were also then given an example shock of 45 V as a demonstration. After this, they were seated behind a wall so that they could hear but not see the stooge who was attached to the machine. Although the stooge 'learner' at no point in the procedure received any kind of shock, the elaborate machine was set up to convince the participants that they were really able to injure the learner. The experimenter remained with the participant; the same experimenter was used in each trial. He was a 31-year-old teacher who wore a grey technician's coat and had a stern manner throughout.

The participants were instructed in the 'memory' task, which involved reading pairs of words aloud to the learner, and subsequently testing the learner on their recognition of the words. Whenever the learner made a mistake, the participants were told by the experimenter to give him a shock by pressing a switch on the generator. They were ordered to increase the level of shock each time by 15 V for each error the learner made. Since the learner was a stooge, they could follow a pre-set plan of mistakes, deliberately giving the wrong answers at particular times.

Until 300 V were reached, the learner had remained silent when receiving the punishment. However, once the punishment level had reached 300 V, the learner began to pound on the wall in protest to the participant. After this time, the learner made no further noises and stopped responding to the memory task altogether. If and when the participants asked the experimenter what they should do, the experimenter insisted that they continue with reading the words aloud and punish the learner, treating no response as an incorrect answer. When participants protested at this, the experimenter continued to give them verbal prods in the sequence: Please go on / Please continue / The experiment requires that you continue / It is absolutely essential that you continue / You have no other choice, you *must* go on. These verbal prods or orders had a set wording, and were given in a standard order to any participants who protested at the task.

Reflections: What do you think the real participants might have been thinking when the learner stopped pounding the wall and did not give any more answers to the task?

The procedure was considered to be complete when the participant refused to give any more shocks, or when they had given the maximum 450 V available. One-way mirrors were used to record the physical behaviours of the participants, and observers noted any comments that were made. After the procedure was complete, each participant was interviewed and had the deception explained to them fully. As part of the interview, participants were asked to estimate how painful they thought the final 450 V shock was

on a scale of 0–14 ('not at all painful' to 'extremely painful'). They were given the chance to meet the learner again, in order to reassure them that they were not injured and to restore the participant's well-being.

Results

Most participants were convinced that all aspects of the situation were real; that they were delivering electric shocks to another person which were extremely painful. The **mean** estimate of the pain of the 450 V shock was 13.42 out of a maximum 14 (as reported in the debriefing interview), meaning that participants were fully aware that they were causing serious pain.

RESEARCH METHODS

The **mean** is a measure of average calculated by dividing the sum of all participants' scores by the number of participants. In this study, 368 V is the mean average shock administered by participants. The mode is a different measure of average: it represents the most frequently occurring score amongst a group of participants. Using the data in Table 5.1, calculate the mode of the participant's shock scores. What conclusion can you draw from this?

Despite the findings being clear that participants believed the situation was real, participants showed extremely high levels of destructive obedience. The mean voltage given by participants was 368 V. All participants gave at least 300 V, and 65% gave the maximum 450 V shock (see Table 5.1). This is a startling contrast with the 0-3% obedience rate estimated by Milgram's students and colleagues prior to the study.

However, the qualitative data collected in this study revealed that participants showed signs of tension when undertaking the procedure. Observers reported signs of nervousness in participants, which increased as they gave more powerful electric shocks. The participants were also frequently observed to be sweating, shaking and groaning, with 14 out of the 40 men showing signs of nervous laughter or smiling. One participant could not complete the experiment because he went into a violent seizure, presumably as a result of the high level of stress he was experiencing.

ISSUES AND DEBATES

The application of psychology to everyday life: this study shows us how likely it is that we follow the orders of an authority figure, even when we are not comfortable doing so. Think about how this might affect people in the military, or working in hospitals where they have to carry out orders which may conflict with their own consciences or personal beliefs.

Comments made by the participants who protested at the orders given included 'I don't think I can go on with this…I don't think this is very humane', and 'I'm gonna chicken out…I can't do that to a man, I'll hurt his heart'. Nonetheless, the verbal prods given by the experimenter were generally successful in persuading the participants to continue. After the procedures ended, the participants showed visible signs of relief, wiped their faces, sighed and shook their heads. A small minority of participants, however, did not show elevated levels of stress and appeared calm during the procedure.

Voltage label	Voltage range (v)	Number of participants for whom this was maximum voltage
Slight shock	15–60	0
Moderate shock	75–120	0
Strong shock	135–180	0
Very strong shock	195–240	0
Intense shock	255–300	5
Extreme intensity shock	315–360	8
Danger: severe shock	375–420	1
XXX	435–450	26

Table 5.1 Distribution of participants' maximum shocks

Conclusions

Milgram's study supports the idea of a situational explanation for obedience. He identifies a number of factors which contributed to the high level of obedience recorded in his study. One of these factors is the perceived legitimacy of the study; the professional academic environment of the study and use of uniform by the experimenter. Other important factors which may encourage obedience include the feeling of financial obligation the participants had towards the experiment, and their belief that both they and the learner had freely volunteered to participate.

Milgram went on to draw two main conclusions from this study:

- Individuals are much more obedient to authority than we might reasonably expect. This seems to be true for the majority of people.
- Despite high levels of obedience, people find the experience of carrying out destructive acts under the orders of authority figures triggers feelings of stress. This is due to a conflict between two important social phenomena: the need to obey those in authority and the need to avoid harming other people.

Strengths and weaknesses

The method used in this study was a controlled observation. This means that it was possible to control extraneous variables in the environment, such as the age and appearance of the actor playing the stooge. This meant that the level of shock administered by each participant was not based on whether the participant felt more or less sympathetic towards different stooges (e.g. they might have been less willing to deliver shocks to an older individual). Also, the procedure was standardised throughout; the verbal prods used by the experimenter were the same each time. The level of control and standardisation of the procedure means the research was more **reliable**, because each participant went through exactly the same experience. The clever design of the electric shock generator and the example shock given to participants improved the **validity** of the design, because it ensured that participants were convinced that the study was real and that their actions actually mattered.

Reflections: Why is validity so important in this study? Why was it so essential participants believed that they were delivering real shocks? Consider the evidence in Milgram's results which support the idea that the study was believable.

The participants in this study were all males, and came from the same local area. This could mean that the sample was low in **generalisability**; it would not be possible to predict what differences there might be in obedience levels between men and women, for example. However, Milgram carefully selected participants to ensure a range of ages and backgrounds. This means that the sample has greater **validity**. It showed that even those with professional backgrounds who are more likely to be in positions of power are susceptible to obeying the commands of an authority figure.

The main measure of obedience was through the voltage of shocks delivered. This is a quantitative measurement, which offered an **objective** record of obedience for each participant. It made it easy to compare the results of the participants and draw conclusions about the overall amount of destructive obedience seen in this study. However, this measure alone did not fully explain the experience of the participants. Qualitative measures such as the notes of observers were used to capture the physical and verbal behaviour of those administering the shocks. Although this data is more **subjective**, it provided a richer understanding of the tension between wanting to obey orders and wanting to obey one's own conscience. Furthermore, interviews with the participants after the procedure also helped explain some of the behaviour of the participants.

This study had major **ethical issues**. Although participants had consented to take part in the research, they did not give their informed consent as they were told the study was about memory and punishment. Participants were repeatedly deceived throughout the study, as well. For example, they were led to believe they had chosen the role of teacher by chance when in fact it was a set-up. Furthermore, participants were arguably denied their right to withdraw. Although they were told they could keep the payment for participation no matter when they left, many felt as though they had to keep going with the shocks out of obligation to the research and as a result of the verbal prods.

ISSUES AND DEBATES

Ethical issues are particularly important to Milgram's research, which caused outrage at the time of its publication. In some ways this research can be considered harmful to the reputation of psychology. It tells us uncomfortable truths about the power of situational factors over ordinary individuals. It might also lead to distrust by the general public who may not want to take part in future research studies.

Participants were not protected from psychological harm; many underwent visible and extreme distress, yet in only one trial was the procedure stopped. All participants were debriefed and told the true aim of the study, as well as being reassured that they had not done any real harm. However, there is the potential for lasting negative consequences to the participants, who may have felt deeply disturbed by their own behaviour.

SELF-ASSESSMENT QUESTIONS

1 The study by Milgram collected both quantitative and qualitative data. Give **one** quantitative and **one** qualitative finding.

2 Milgram's findings support the situational explanation for obedience. Identify **two** features of the situation which may have contributed to the high levels of obedience seen in this study.

5.2 Core study 2: Piliavin et al. (subway Samaritans)

Piliavin, I. M., Rodin, J., & Piliavin, J. (1969). Good Samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology, 13*(4), 289–299.

Aim

Following on from previous laboratory-based studies, the researchers aimed to study **bystander** behaviour in a natural setting. They also wanted to investigate the effect of four situational variables on helping behaviour or '**Good Samaritanism**':

- the type of victim
- the race of the victim
- the behaviour of a 'model'
- the size of the group of bystanders.

KEY TERMS

bystander: a person who is present at, but may not be directly involved in, a particular situation. 'Bystander apathy' or the 'bystander effect' refers to the actions of bystanders who don't help others in the event of an emergency.

Good Samaritan: this term originates from the New Testament in the Bible. It refers to a story of a Samaritan (person originating from ancient Samaria) who stops to offer help to an injured stranger.

Background

A key trigger for research into bystander behaviour was the murder of a young woman called Kitty Genovese (Figure 5.4) in New York City in 1964. After returning from her work during the early hours of the morning, Miss Genovese was followed and assaulted by an assailant near her home. One witness had called down to warn off her attacker as Miss Genovese screamed that she was being stabbed. Her attacker was scared off but returned to continue the assault. It was alleged in news reports of the time that around 38 individuals living nearby were either eye or ear witnesses to the crime, but failed to prevent her murder. While it is impossible to say what each person saw or interpreted, the event triggered the interest of social psychologists who tried to understand the behaviour of the bystanders.

ALL THE CITY'S NEWS — DAILY NEWS, NEW YORK'S HOMETOWN NEWSPAPER

Kin see Kitty's killer

THE KITTY GENOVESE FILE

PAST This is how the Daily News covered the famous murder case in 1964.

Crowd Applauds As Queens Jury Dooms Moseley

Knifed Barmaid Dies in Mystery

A Nighttime Urge to Slay: Held in Killing of 2 Women

PRESENT Her brothers Vincent (left) and William showed up at Brooklyn Federal Court to confront their sister's killer.

He asks new trial amid their glares

Family sees nightmare in the flesh

MICHAEL DALY

5.4 Newspaper article about Kitty Genovese

Reflections: With so many witnesses to the murder of Kitty Genovese, stop and ask yourself why people might have been reluctant to get involved. Did they just not care, or could there be other explanations? Try to think of as many explanations as you can.

One explanation for the lack of bystander helping was outlined by Darley and Latané (1968). They found that bystanders who believed that there were other people witnessing an emergency, such as over hearing someone having an epileptic seizure, were significantly less likely to help than those who believed they were alone in hearing the event. The explanation for this is known as the **diffusion of responsibility** hypothesis. Alternatively, if we witness those around us assisting or 'modelling' helping behaviour, we may be more likely to imitate and engage in helping.

There is also evidence to suggest that we are more likely to help some people than others. For example, we may be more willing to assist those whom we perceive as more similar to ourselves. Some studies have shown that bystander helping occurs more when victims are not seen as responsible for their circumstances (Schopler and Matthews, 1965). This may be because they evoke less sympathy from those around them.

KEY TERM

diffusion of responsibility: a person is less likely to take action in an emergency where there are others there also able to help. In a large group, the perceived sense of individual responsibility towards those in need is 'diffused' or reduced to the extent that people feel little obligation to intervene. An explanation for the bystander effect.

Method

Research method and design

This study was a **field experiment**. This means it took place in a realistic environment; in this case, the New York City subway. It can be described as using an independent groups design, as the trials were repeated on different days and involved different participants in each condition.

Reflections: The study intended to use different participants for each of the trials. Why might this not have been guaranteed? Consider the impact this might have had on the results of the study.

There were four independent variables (IVs) which corresponded to the factors outlined in the aims of the study. They were operationalised as:

- the type of victim: the levels were 'drunk' or 'ill' victim
- the race of the victim: the levels were black or white victim
- the behaviour of a 'model': the levels were a model who was either close to or distant from the victim helped, either early or late in the event
- the size of the group of bystanders: this level was the naturally occurring number of passengers present in the subway carriage.

The dependent variable (DV) was the level of bystander helping. In quantitative terms, this was operationalised as the time taken for the first passenger to help, as well as the total number of passengers who helped. The race, gender and location in the carriage of each helper were also recorded. Qualitative data was recorded in the form of verbal remarks made by passengers during each incident.

RESEARCH METHODS

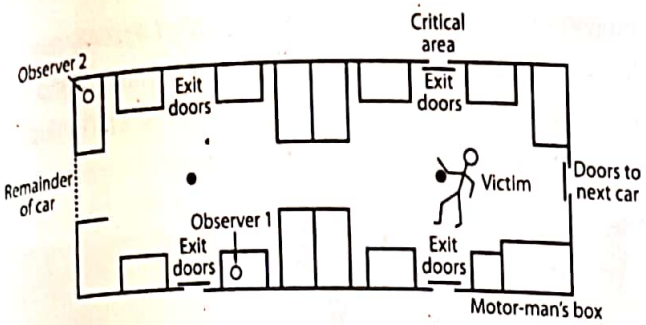
A **field experiment** is a type of study that takes place in everyday locations, rather than the controlled environment of a lab. It still has an experimental design, meaning that independent variables are manipulated by the researcher, while the dependent variables are measured.

Sample

This study took place on the New York subway. Participants were passengers travelling on an underground service between Harlem and the Bronx weekdays between 11 a.m. and 3 p.m. We might regard this 'unsolicited' sample as an opportunity sample as they were not deliberately selected for participation. The total estimated number of participants was 4450 people, of whom around 45% were black and 55% were white. The mean number of passengers per carriage was 43, and the mean number of people in the critical area (where the incident took place) was 8.5.

Procedure

Four teams of student researchers carried out the study, following a standard procedure. On each trial, two male and two female students boarded the train using different doors. The female confederates sat in the area adjacent to the immediate 'critical' area where the incident took place (see Figure 5.5). They observed the passengers and recorded data during each trial. The male confederates took the roles of the victim and the model. The victim stood at the pole in the centre of the critical area, and the model remained standing throughout the trial.



5.5 Layout of adjacent and critical areas of the subway carriage

Each trial used the same route, because it included a 7.5 minute gap between two stations. At approximately 70 seconds into the journey, the 'victim' staggered forward and collapsed. He remained lying on the floor looking upwards. If he received no help, the model would help him to his feet at the next stop.

RESEARCH METHODS

Even though this is a field experiment, there was considerable **standardisation** between trials and **controls** that ensured there were few differences between different conditions. For example, the 'victim' was always identically dressed and behaved in the same way for all trials.

The victim was played by different males during the study, but all were made to look similar. They were aged 26 to 35 years; three were white and one was black. They were dressed in identical, casual clothing (jacket, old trousers, no tie). On 38 out of 103 trials the victim smelled of alcohol and carried a bottle of alcohol wrapped in a brown bag. On the remaining 65 trials they appeared sober and carried a black cane. In all other ways they behaved identically.

Reflections: There were an uneven number of 'drunk' and 'cane' trials. The student confederates reported that they did not like playing the drunk victim. Why might this be?

The models were all white males aged 24 to 29 years of age, who were also dressed informally. When helping, the model raised the victim to the sitting position and stayed with him until the train reached the next stop.

Trials were split into the following conditions:

- Critical/early: model stood in critical area and waited 70 seconds to help victim
- Critical/late: model stood in critical area and waited 150 seconds to help victim
- Adjacent/early: model stood in adjacent area and waited 70 seconds to help victim
- Adjacent/late: model stood in adjacent area and waited 150 seconds to help victim
- No model condition: the model did not help the victim until after the trial was over and the train had reached the next stop.

Results

Overall, the frequency of helping recorded in this study was much higher than had previously been reported in laboratory studies. The majority of the helpers were male. Nearly 80% of victims received spontaneous help (i.e. helped before model intervened or in a no-model condition), and in around 60% of cases more than one person helped.

There were key differences in levels of helping between different conditions of the study, as seen in Table 5.2. In terms of the type of victim, participants were more likely to help the victim with the cane than the drunk victim (the cane victim received help in 62/65 trials; the drunk victim received help in 19/38 trials). In the cane trials, spontaneous helping also occurred earlier than in the drunk trials. For example, in all but three of the cane trials that were also model trials, helping occurred before the model could give assistance.

Trial	White victim		Black victim	
	Cane	Drunk	Cane	Drunk
No model	100%	100%	100%	73%
Model trial	100%	77%	–*	67%

*No model trials for the black 'victim' were run for the cane condition

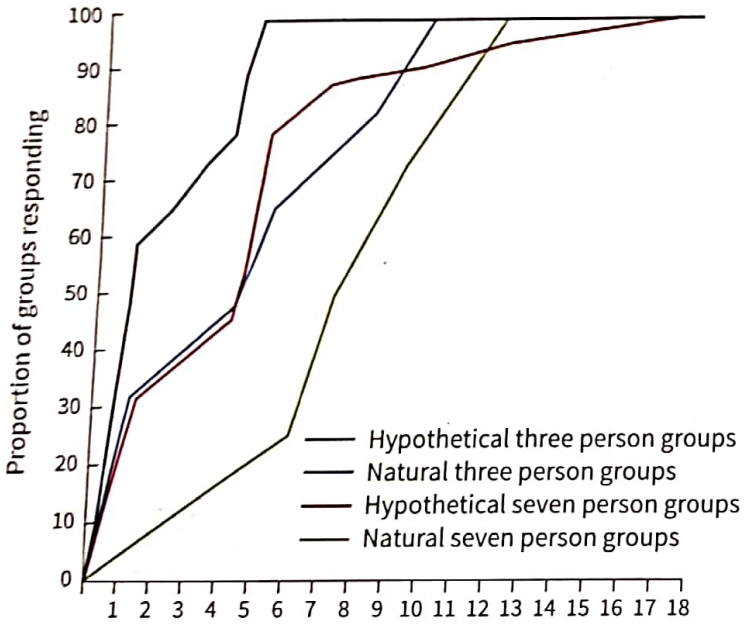
Table 5.2 Percentage of trials in which help was given

In terms of race, both black and white cane victims were equally likely to receive help. However, there was some minor evidence of same-race helping in the drunk condition, with participants being more willing to offer help to those of their own race. In the drunk condition, black victims were found to receive less help overall. Although these results were non-significant, they would have supported research suggesting people are more likely to help those similar to themselves, as they feel more empathy towards them.

The effect of modelling was difficult to analyse, because most of the helping that occurred was spontaneous. However, it appeared that early model intervention at 70 seconds was slightly more likely to result in helping behaviour than waiting until 150 seconds had passed.

Reflections: Piliavin et al. found that people are more likely to help when they had watched a model help quite quickly after the victim collapsed. What explanation might there be for this?

Surprisingly, unlike previous studies, this research found no evidence to support the diffusion of responsibility hypothesis. In fact, there was some evidence to suggest that when more passengers were present, rates of helping were also slightly higher. Looking at the graph in Figure 5.6 we can see that the hypothetical speed to respond for seven-person groups, as predicted by the diffusion of responsibility theory, is slower than for three-person groups. This is because in seven-person groups the responsibility should be diffused or shared between more individuals. In fact, natural seven-person groups were faster to respond than predicted, and faster to respond than the three-person groups. This directly opposes the prediction of diffusion of responsibility.



5.6 Graph showing the difference in helping behaviour between the groups

Additional observations showed that the majority of helpers were male. Observers noted that in around 20% of trials, passengers actually moved away from the critical area where the incident was taking place. There were a higher number of comments made during trials without helping. There were also more comments made during trials with a drunk victim.

Piliavin et al. proposed an alternative explanation for their findings, known as the '**cost–benefit model**' (Figure 5.7). They suggest that witnessing an emergency raises an individual's level of arousal. They may either become more likely to feel disgust and aversion, or even sympathy and courage. This heightened arousal level prompts individuals to act, in order to reduce difficult feelings.

KEY TERM

cost–benefit model: involves a decision-making process in which a person weighs up both the advantages and disadvantages of helping. If it seems beneficial to help, then the person is more likely to do so; if the risks are too great, they may refrain

5.7 Weighing up the pros and cons of helping

Reflections: Using the cost–benefit model proposed by Piliavin et al., consider the costs and benefits of helping and not helping in both the cane and drunk conditions. Does this support the findings of the study?

Potential responses include helping directly (as seen in most trials in this study) or indirectly (asking others to help). Individuals may alternatively leave the area where the incident is taking place (as seen in some trials) or determine the victim should not be helped. This final response is evidenced in comments by passengers who expressed disgust or dismissal at the drunk victim's collapse.

Conclusions

This study found that in a natural setting, many people would offer spontaneous help to a stranger, even in a group situation. This study found no evidence of diffusion of responsibility, but did identify several factors which may determine decisions to help:

- the type of victim (someone using a cane will be helped more than a drunk person)
- the gender of the helper (men are more likely to help than women)

- people may be more likely to help members of their own race, especially if the victim is drunk
- the longer an emergency continues, the less likely it is that anyone will help, and the more likely it is they will find another way of coping with arousal.

ISSUES AND DEBATES

Participants in this study were not aware they were taking part in research. This means they did not give their informed consent to participate. Why might this be an **ethical issue**, and how could it be overcome?

Strengths and weaknesses

The method used in this study was a field experiment which used independent measures. This meant that it had good **ecological validity**; the participants were ordinary train passengers who were unaware they were taking part in the experiment. They would have behaved naturally as they believed the emergency situation to be real. However, one limitation of this method is that there is less control over extraneous variables, such as the weather conditions or train delays, which could affect the participants' behaviour and lower the **validity** and **reliability** of the study.

There are other methodological issues with the study. For example, the experimenters cannot be sure that participants only took part in the experiment once; as they used the same route each time there is a chance participants may have been exposed to more than one condition of the experiment. Suspecting that the emergency was a set-up might have made the participants more or less likely to offer help, creating **demand characteristics**.

Reflections: Field experiments are often affected by a lack of control of extraneous variables. Can you identify ways in which the researchers improved validity through standardising their procedure?

The participants in this study were all subway passengers from New York City, which means that the **sample** is unrepresentative. It would not be possible to predict levels of bystander helping in other countries from this study. However, the design of the study meant that around 4500 individuals participated in the study, which included a mix of ethnicities and genders. This large sample therefore is likely to be quite representative and has greater **validity**.

The main recorded measure of bystander helping was the number of helpers and how long they took to help. This quantitative measurement ensured an **objective** record, made more **reliable** by the presence of two observers. The observers also recorded qualitative data including the remarks and movements made by the passengers during each trial. This allowed the researchers to understand the thoughts and behaviours associated with helping in more depth.

This study raised serious **ethical issues**. Firstly, participants did not give their consent to take part in the research, nor were they debriefed after the study had finished. Participants were deceived during the study, as they believed the victim had genuinely collapsed and needed help. They might have suffered psychological distress as a result of the study, guilt at not helping or concern about the well-being of the victim.

SELF-ASSESSMENT QUESTIONS

3 The aim of Piliavin et al.'s study was to test the diffusion of responsibility hypothesis. Outline this hypothesis in your own words.

4 Piliavin et al. carried out a field experiment. Identify **one** strength and **one** weakness of this research method.

5.3 Core study 3: Yamamoto et al. (chimpanzee helping)

Yamamoto, S., Humle, T., & Tanaka, M. (2012). Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. *Proceedings of the National Academy of Sciences, 109*(9), 3588–3592.

Aim

The aim of the study was to learn more about helping behaviour in chimpanzees. Specifically, the researchers wanted to find out:

- whether chimpanzees can understand the needs of **conspecifics**
- whether chimpanzees can respond to those needs with targeted helping.

KEY TERM

conspecific: member of the same species. Therefore, in this study, 'conspecifics' refers to other chimpanzees.

Background

In order to maintain cooperative societies, humans engage extensively in helping behaviours (Figure 5.8). Other animals engage in helping, but more often at the request of their conspecifics. In other words, they may not help spontaneously or voluntarily but rather at the direct request of others. The ability to offer targeted help to members of our own species relies on an understanding of their goals, which is linked to 'theory of mind' (ToM) ability.

5.8 Altruistic helping in human society

As you may remember from Section 3.2, ToM refers to our capacity to understand the intentions and needs of others. Some people believe that ToM is unique to humans, and can explain why they are the only species to demonstrate **altruistic** helping. However, some recent studies have shown that some primates have the capacity for helping and food sharing, without direct benefit to themselves (Figure 5.9)

KEY TERM

altruistic: acting helpfully towards others without obvious benefit to oneself. An example of this might be chasing after a stranger in order to return their wallet, which involves no guarantee of reward.

5.9 Direct requests for help elicit more responses in chimpanzees

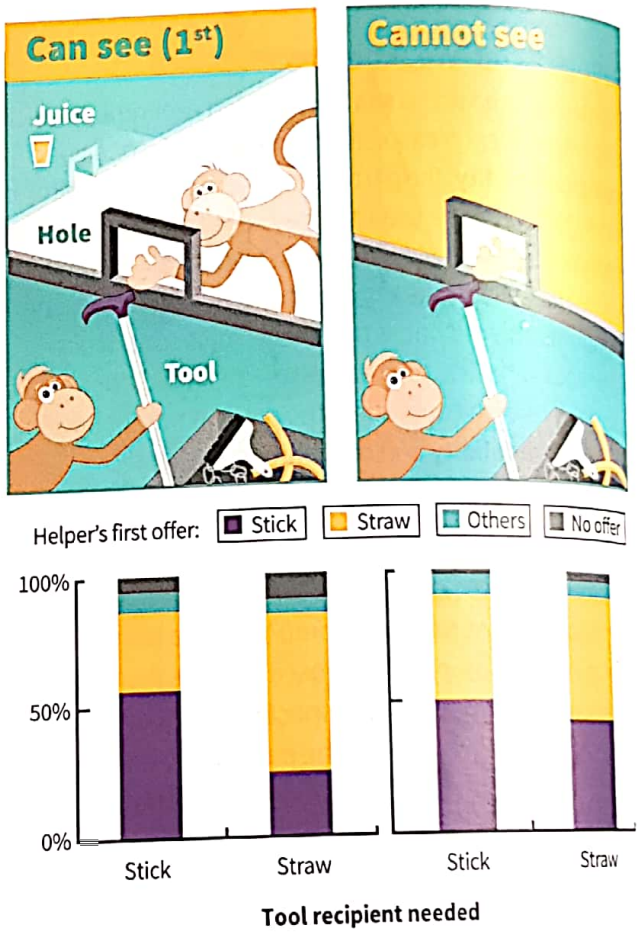
One explanation for conspecific animal helping is known as targeted or instrumental helping. It is a cognitive explanation for social behaviour; where the type of care or help given is based on a cognitive understanding of the need or situation of others. There is some evidence to suggest that chimpanzees can engage in targeted helping following direct requests (e.g. an outstretched arm requesting food), but little is known about whether they have any ability to interpret the needs of conspecifics.

Reflections: Can you think of some everyday examples of targeted helping? Try to think of three different situations. Consider how we might guess that someone needs our help, even if they don't directly ask for it.

Method

Research method and design

This study was a laboratory experiment. It took place in an artificial environment where the chimpanzees were seated at adjacent experimental booths (see Figure 5.10).



5.10 Illustrations of the 'can see' and 'cannot see' condition

The independent variable (IV) was the ability of the chimpanzee to give targeted helping to another chimpanzee. There were two conditions in the task. In the first, the potential helper chimpanzee was able to see the other's tool use situation, in the second they could not see. The study used a repeated measures design, which means that all the chimpanzees took part in both conditions of the experiment.

Reflections: This study used a repeated measures design. Why do you think this design was chosen for this experiment in preference to an independent measures design?

The dependent variable was the targeted helping behaviour. This was operationalised as the items offered by the participants to conspecifics. The item offered was either the correct tool (stick or straw, depending on situation) or an incorrect non-tool item (e.g. a piece of string). The behaviour of the participants was recorded on video camera and was used to produce quantitative data; the number of correctly targeted offers per condition. The video also captured the behaviour of the chimpanzees, such as how they moved, responses to gestures, as well as where they were looking while they sat in the experimental booths.

Sample

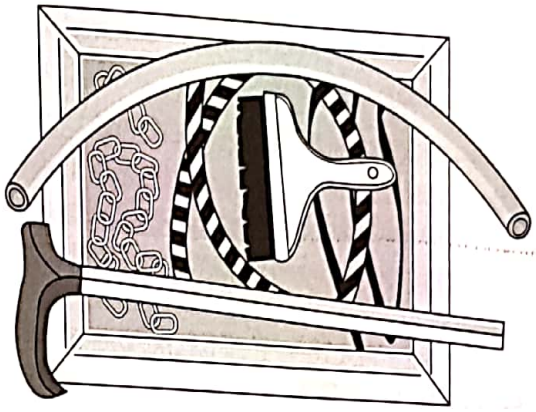
Five chimpanzee participants who were socially housed within the Primate Research Institute at Kyoto University took part in this study. Each had previously been a part of a number of other perceptual and cognitive studies, including some investigating helping behaviour in a similar setting to the present study. In each trial, the chimpanzee participants were paired in mother and child pairs. This was because each pair had already demonstrated frequent tool-giving interactions in previous experiments. Also, they were familiar with the tool-use tasks used in this study.

Reflections: Experiments with primates often rely on the use of charts composed of symbols. In this study, the chimpanzees communicated directly with one another. This meant that their natural communicative abilities could be observed in situations of flexible and spontaneous helping. What do you see as the challenges of directly observing chimpanzee communication?

Procedure

The experimenters designed the experimental task in order to examine the ability and flexibility of chimpanzees to help a conspecific, depending on their need. A chimpanzee had to select a tool that would help the other chimpanzee to solve a problem. One task required a stick and the other required a straw. Solving the task allowed the second chimpanzee in the pair to obtain a reward: a juice drink.

Seven objects (including a stick and a straw) were placed in a booth occupied by a potential helper (see Figure 5.11). This could not be reached by the potential recipient but could be requested by the chimpanzee poking his or her arm through a hole to gesture. This allowed the experimenters to examine whether the potential helper chimpanzee was able to understand what the other needed. Before any trials started, the chimpanzees went through a familiarisation phase each day where they could examine and manipulate all the objects without the need to use them as tools or offer them to others.



5.11 Tray of items presented to participants

When the experimental trials began, each chimpanzee experienced the conditions in the same order. Firstly they were placed in the 'can see' booth in which the panel between the two chimpanzees was transparent. Next they completed the task in the 'cannot see' booth, in which the panel was opaque. Finally the 'can see' condition was repeated in order to confirm that any difference in object choice between the first two conditions was due to intentional, targeted helping and not an **order effect**.

RESEARCH METHODS

An **order effect** is when the order of conditions in an experiment has an effect on participants' behaviour. This can be a confounding variable; meaning that the validity of the results can be lowered because participants are tired, bored or have worked out what they are supposed to be doing in the study. How else could the experimenters in this study avoid creating an order effect?

Forty-eight trials were carried out in each condition; this consisted of a random order of 24 stick-use and 24 straw-use situations. Trials began when the tray of objects was presented. The trial ended when the recipient received the object and succeeded in obtaining the juice reward or after five minutes had elapsed without an object being passed. 'Offers' were counted when the chimpanzee held out the object to the recipient, whether the recipient took the object or not. Only the first offer of help was counted; subsequent offers of different items were not included in the data. Between two and four trials were conducted per day.

Results

The experimenters found that chimpanzees are capable of flexible targeted helping based on an understanding of

the other's goals. In the 'can see' condition, objects were offered in 91% of trials, and this mostly occurred following requests from the paired chimpanzee (i.e. holding out hand through the hole in the panel). This can be compared to the pre-test familiarisation trials, where 'offering' occurred in only around 5% of trials.

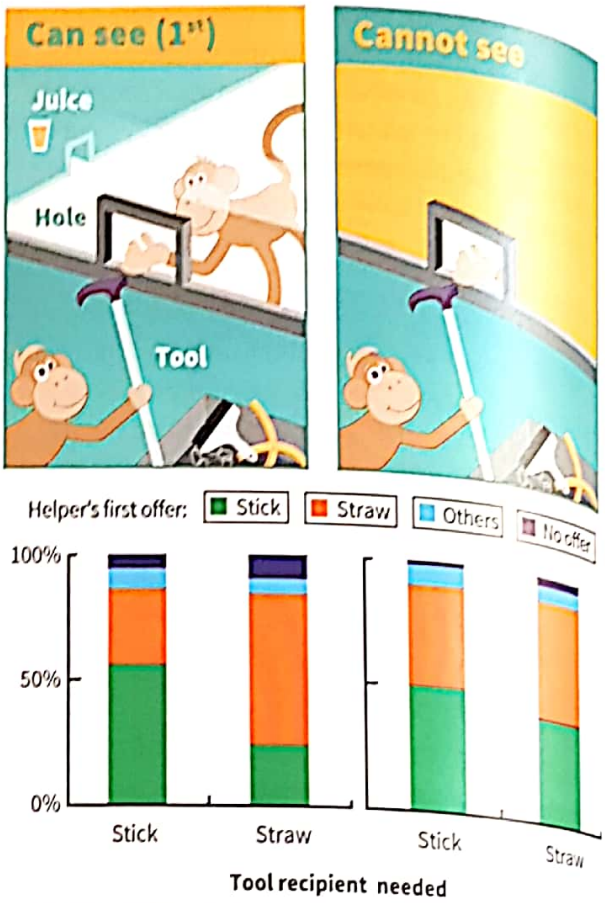
Apart from one individual, all chimpanzees first offered tools (stick or straw) on significantly more occasions than non-tool objects. This bias suggests that the chimpanzees were able to discriminate between potential tools and non-tools. In addition, the chimpanzees selected the correct tool (stick or straw) to offer their partner over a significant number of trials. This suggests the chimpanzees used targeted helping through understanding of the task confronting their partner.

In the 'cannot see' condition, at least one object was offered in 96% of trials. Again, offering occurred mostly at the request of the paired conspecific. Like the first condition, all but one chimpanzee first offered a potential tool significantly more often than non-tools. This chimpanzee, 'Pan', mainly offered the brush item at first. When this item was removed from her tray, she made appropriate tool offers similar to the other chimpanzees.

An important difference was observed between 'can' and 'cannot see' conditions: in the 'can see' condition, there was a significant difference in which tool (stick or straw) was offered, depending on the task requirement (Figure 5.12). In the 'cannot see' condition there was no significant difference in which tool was offered in all but one of the chimpanzees. The one chimpanzee named Ayumu who did select the correct tool more often was able to stand and peer through a hole in the wall. He did so in order to view his partner chimpanzee (his mother) and observe the task that she was facing.

The results suggest that the chimpanzees were only able to understand their partner's goals when able to see the task themselves. The partner chimpanzees performed request actions with similar method and frequency in both conditions; this behaviour did not communicate information regarding specific needs (such as the appropriate tool required).

The third trial, a repeat of the 'can see' condition, was undertaken with three chimpanzees who had previously shown a significant difference in tool selection in the 'can see' condition and a non-significant difference in the 'cannot see' condition. Object offer was observed in 98% of trials; offer upon request accounted for around 80%. There was a significant difference in offer of stick or straw, depending on the partner chimpanzee's situation. This confirms flexible targeted helping with an understanding of the tool needed to complete the task, when the chimpanzees could see the task for themselves.



5.12 Results of the helping conditions

Conclusions

Chimpanzees can understand the needs of conspecifics in order to help them in successfully solving tasks.

- They will offer help to conspecifics who require it in the majority of cases, but usually as a response to a direct request rather than as a spontaneous act.
- Chimpanzees rely on visual confirmation of conspecifics' needs in order to offer targeted helping.

Strengths and weaknesses

The method used in this study was a laboratory experiment. There were high levels of control in the study and a standardised procedure. For example, the presentation of the objects on the tray was the same for each trial, and the chimpanzees sat at the same booths to undergo each trial. These measures increased the **reliability** of the study. Furthermore, the experiment used a repeated measures design. This design meant that the

chimpanzees participated in both or all the conditions of the study; reducing any risk of individual differences and increasing **validity**.

The study had low **ecological validity** because it was conducted in an artificial environment. The chimpanzees were given tasks and tools that they would not normally use in their natural environment. However, as the chimpanzees had previously taken part in laboratory studies and were therefore familiar with the task and materials used in this research, they probably showed normal behaviour. In this way it could be argued that the study was **valid**.

On the other hand, concluding that offering the correct tool is an intentional cognitive decision that involves theory of mind may not be an accurate assumption. It could be argued that the findings of this study could be a result of an automatic association created by previous experience. For example, Pan repeatedly offered the brush to her partner in early trials. This might suggested that her previous experiences with similar tasks created **bias** in her responses to this study.

The participants in this study were five chimpanzees, and came from the same research institute. This means it was a very small **sample** that is arguably low in **generalisability**. It would be difficult to say this sample of captive chimpanzees is representative of wild chimpanzee populations, although there is nothing to suggest that any of the animals was unique or unusual.

RESEARCH METHODS

You might think that only humans respond to **demand characteristics** but there is a range of evidence that suggests some animals do too. By looking over the top of the opaque panel, and peeking through the hole, chimpanzees in this study made it clear that they understood the nature of the task. This suggests they had an awareness of the demands of the task.

Targeted helping was observed using video recording and quantified in a standard way as no offer, offer of tool or offer of other item. Both the way in which the data was recorded and the type of data recorded provide an **objective** record of helping for each participant. Other qualitative data was gathered during the experiment such as the behaviour of one chimpanzee who looked over the opaque panel in the 'cannot see' condition. This data is important in helping us understand why the chimpanzee then showed an increase in correctly targeted helping.

Reflections: No physical harm was caused to the chimpanzees during this study. What other types of harm should psychologists consider when conducting research with animals? Do you think any of these issues apply to the study by Yamamoto et al.?

This study involves animals and has some specific **ethical issues**. Although issues of consent and right to withdraw do not apply to research of this kind, there are important considerations for the animals' well-being. The researchers state that the study was approved by the Animal Care Committee of the Primate Research Institute at Kyoto University. Chimpanzees were tested and cared for in accordance with the guide produced by this committee. While specific details are not given in this study, guidance of this type typically recommends that treatment of animal participants is humane. This might involve ensuring that the basic needs of animals are met, such as adequate access to food and shelter (Figure 5.13). This study focuses on helping tasks, and does not involve punishment or physical harm to the chimpanzees.

5.13 Chimpanzees in captivity

SELF-ASSESSMENT QUESTIONS

5 The research by Yamamoto et al. used experimental booths in their set-up. Give **two** details describing the set-up of the booths.

6 This study used a repeated measures design. Identify both conditions of the IV which all participants completed.

5.4 Issues, debates and approaches

Application of psychology to everyday life

Milgram's study has important implications for understanding obedience in the real world. Previously it was believed that acts of extreme, destructive obedience such as the Holocaust had a dispositional explanation (i.e. that Germans were somehow very different from other people). However, Milgram's later research (1974) and other work inspired by his findings has shown that situational variables such as the legitimacy of authority can elicit destructive obedience in nearly anyone. A result of this is a raised awareness of the power of authority in the workplace. For example, some hospitals have introduced whistle-blowing policies to encourage reporting of mistakes by doctors or other senior staff in order to protect the safety and well-being of patients.

Piliavin et al.'s study has interesting practical applications as well. It tells us about specific situational factors which may make bystanders more likely to help. For example, it may be useful to know that people may be more willing to help if they are of the same sex or race as the victim, or in a situation that they cannot easily just walk past. It shows us that a person in need may be more likely to get help from a stranger if they do not risk embarrassing, intimidating or disgusting them.

The study by Yamamoto et al. demonstrates that chimpanzees, like humans, have the capacity and inclination to help conspecifics. It can also help us to understand more about how chimpanzee societies work in the wild.

Individual and situational explanations

Milgram's research was particularly significant in highlighting the importance of situational factors in influencing how obedient we are to authority figures. Findings from his study showed that the majority of people will be destructively obedient if they feel that the authority figure is legitimate or the cause is worthy, for example.

However, some individuals are more resistant to authority than others – not all participants inflicted the full voltage on the learner as directed. This suggests that there is a role for individual factors, which may affect overall levels of obedience as well.

Reflections: Piliavin et al. also considered situational factors. Can you explain how each of the trial conditions (victim type, race, model, size of group) influenced helping behaviour?

Yamamoto et al.'s study is interesting in that it shows the influence of individual and situational factors on the helping behaviour of chimpanzees. Most chimpanzees demonstrated similar patterns of behaviour in a social situation because they offered tools to conspecifics in need. However, one chimpanzee showed a clear preference for a specific non-tool, possibly as a result of their own prior learning. Another chimpanzee initiated problem-solving in the 'cannot see' condition by peering over the opaque panel to observe his partner. Both of these cases demonstrate the influence of individual personalities on common social behaviours.

Use of animals in psychological research

The study of flexible, targeted helping by Yamamoto et al. is useful in that it can allow us to make some comparisons between chimpanzees and humans. For example, a more recent study has shown that such chimpanzee helping approximates the level of helping shown by 18-month-old toddlers (Svetlova et al., 2010). They have capacity for understanding the needs of others, as well as a willingness to help. In contrast, by around 30 months, humans have developed capacity for spontaneous helping. However, the cognitive processes involved in theory of mind in humans and primates are not yet well understood. Furthermore, the culture of chimpanzee societies is quite different from humans (e.g. family structures, group sizes and hierarchies) which makes direct comparison difficult.

Summary

Milgram's study investigated the extent to which ordinary people obey the orders of an authority figure when they involve physically harming an innocent person. All participants were willing to give high levels of electric shocks at the command of a researcher, with a significant majority of people willing to give the maximum shock level possible. This was a well-controlled naturalistic observation collecting measure of the dependent variable through objective measurements which were reliable. Qualitative data suggested that most participants showed signs of stress during the destructive obedience. The findings suggest that situational factors may be better at explaining obedience to authority than dispositional ones.

Piliavin et al.'s study considered different factors which affect bystander behaviour. It looked at how specific circumstances might make subway passengers more or less likely to help a man who had collapsed on a train. This was a field experiment involving a large number of participants. Quantitative measurements showed that victims who were ill were more likely to receive help than those who appeared drunk, and that the number of bystanders does not affect the amount of bystander helping. The findings suggest that diffusion of responsibility is not typical of bystander helping in natural environments.

Yamamoto et al.'s study looked at flexible targeted helping in chimpanzees. As a laboratory experiment it was well controlled and was able to show that chimpanzees can offer carefully targeted help to conspecifics. Chimpanzees were able to offer tools to their partners in most cases, and usually only when assistance was directly requested. Careful observation of the participants' behaviour provided reliable measures of helping, although the experiment only used a small sample. The findings show that humans may not be unique in possessing some capacity for theory of mind as they suggest chimpanzees can understand the goals and needs of others.

Exam-style questions

1 Although Milgram's study (obedience) is often used as an example of poor ethics, there were many ways in which he made good ethical decisions.
   a Describe **two** guidelines which Milgram could be said to have tried to follow. **[2 marks]**
   b Explain how **each** of these two ethical guidelines was followed by Milgram. **[4 marks]**

2 Piliavin et al. (subway Samaritans) concluded that their results did not support the 'diffusion of responsibility' hypothesis. Explain this conclusion using evidence from the study. **[6 marks]**

3 State **two** aims from the study by Yamamoto et al. (chimpanzee helping) and explain whether each one was supported by the study. **[6 marks]**